

DESIGNING TOOLS FOR COLLABORATIVE SENSEMAKING
DURING COMPLEX CRIME ANALYSIS

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Nitesh Goyal

August, 2017

© 2017 Nitesh Goyal

DESIGNING TOOLS FOR COLLABORATIVE SENSEMAKING DURING COMPLEX CRIME ANALYSIS

Nitesh Goyal, Ph. D.

Cornell University 2017

As data grows complex, making sense of complex data for problem-solving by teams is becoming a challenge. Previous research suggests that challenging data problems can only be solved by leveraging human cognition, in combination with computational advances. However, this potential remains untapped, as collaborative sensemaking is fraught with significant multiple socio-cognitive challenges of information sharing and analysis. Lack of human centered design and evaluation approach to develop information sharing and problem-solving tools have resulted in little empirical knowledge about these challenges and potential design solutions to overcome these challenges.

This dissertation offers a human centered design approach to iteratively design and evaluate collaborative sensemaking tools for a problem-solving task in the crime-solving domain. Crime-solving domain offers a life-critical germane ground to design for known human challenges that continue to recur. For every challenge, I designed and deployed a tool, and evaluated its effectiveness in a laboratory experiment where participants used my tools to solve a crime problem collaboratively and reported on task-performance and collaboration experience. Tools deployed in each iteration benefitted from the objective results, self-reported perception, user-log analysis, video-analysis, and qualitative feedback, from the previous iteration.

First, I designed SAVANT as a modular tool to highlight that data analytic tools perform better when customized for simplicity to enable different sensemaking tasks at hand, as opposed to offering all the complex features always. Next, SAVANT was modified based on lab-experiment to solve the social challenge of inefficient explicit sharing of information among crime analysts. Collaborative version of SAVANT offers implicit sharing of notes and insights as an alternative between remotely collaborating data analysts. Finally, I invented Sensemaking Translucence, a design metaphor used to overcome the cognitive challenge of biased decision-making through implicit visualization of decision-process artifacts, deployed in REFLECTIVA. By leveraging intermediate data analytic artifacts, including notes, insights, and communication, to drive visualizations in SAVANT and REFLECTIVA, my findings would benefit the design of web based data analytic tools. Future research directions and design implications deriving from these findings are also outlined at the end of this dissertation.

BIOGRAPHICAL SKETCH

Nitesh (Tesh) Goyal completed his Ph.D. in the Department of Information Science at Cornell University in 2017. Prior to Cornell, he completed his MSc. in Software Systems Engineering at Rheinisch-Westfälische Technische Hochschule (RWTH), Aachen, Germany. While in Germany, he received a German Government fellowship as an RWTH Research Ambassador to conduct research for his MSc thesis at Berkeley Institute of Design, University of California, Berkeley. Both at Cornell and Aachen, Tesh has been involved in co-founding safe spaces for minority student communities. Tesh received his Bachelors degree in Information Technology (IT) at Indraprastha University's central campus School of IT, while pursuing research at Technical University, Munich, Germany.

Tesh's primary research interests lie in the areas of human-computer interaction, computer supported cooperative work, artificial intelligence, and educational technology. Beyond computing, his interdisciplinary interests cross communication, design, cognitive science, and social sciences. His work has focused on designing tools to support collaborative sensemaking in complex data analysis by remote team members, leveraging spatial awareness tools for collocated sensemaking, and using physiological sensors and crowdsourcing to improve collaborative sensemaking.

THIS DISSERTATION IS DEDICATED TO MY FAMILY AND TO THE COMMUNITY OF
RESEARCHERS THAT HAVE CONTINUED TO PUSH ME FURTHER

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH	i
TABLE OF CONTENTS	iii
LIST OF FIGURES	v
ACKNOWLEDGEMENTS	vi
CHAPTER 1: INTRODUCTION	1
Specific Research Goals and Approach	3
Research Contributions	4
Outline	5
CHAPTER 2: BACKGROUND	6
The Sensemaking Process	6
Tools for Collaborative Sensemaking	11
This Dissertation: Designing Technology for Sensemaking in Crime-Solving	17
Summary	20
CHAPTER 3: SAVANT: A MODULAR TOOL TO SUPPORT CRIME ANALYSIS	21
The Savant System Design	23
Laboratory Study Comparing Visualization and Notepad Tools	27
Results	32
Discussion	40
Limitations and Further Directions	42
CHAPTER 4: COLLABORATIVE SAVANT: DESIGNING FOR IMPLICIT SHARING OF INFORMATION DURING CRIME ANALYSIS	44
Method	49
Measures	52
Results	56
Discussion	61
Summary	63
CHAPTER 5: REFLECTIVA: DESIGNING FOR SENSEMAKING TRANSLUCENCE	64
Study Hypotheses	68
Method	72
Results	77
Discussion	86
Limitations and Future Directions	89
Summary	91
CHAPTER 6: GENERAL DISCUSSION	92

Summary of Results	93
Design Implications	94
General Limitations	98
Future Work	100
Conclusion	104
REFERENCES	105
APPENDIX A.1	135
APPENDIX A.2	150

LIST OF FIGURES

Figure 1: A screenshot of the SAVANT system	24
Figure 2. The visualization showing filtered results	26
Figure 3. Percentage of participants correctly identifying the serial killer	33
Figure 4. Number of clues recalled (out of 9)	34
Figure 5. Percentage of time spent	36
Figure 6. Mean ratings of tool usefulness	37
Figure 7. The Document Space	46
Figure 8. The Analysis Space	47
Figure 9 Results	56
Figure 10. The REFLECTIVA Document Space	66
Figure 11. The REFLECTIVA Analysis Space	67
Figure 12. Sample sensemaking trajectory of the Sensemaking Translucence interface	71
Figure 13. Serial killer identification and number of correct clues	77
Figure 14. Perception of interface usefulness by interface condition	78
Figure 15. Self-reported workload and team experience by interface condition.	80
Figure 16 The Experimental Setup	101
Figure 17. Workflow for Crowdsourcing	102

ACKNOWLEDGMENTS

I would like to thank Susan R. Fussell for her immense help, support, guidance and advice towards this research, and personal life. I met Sue at an INTERACT Workshop as an SV by a chance and ended up pursuing my research training as her PhD student the next year onwards at Cornell, where this dissertation was written. Through these multiple years, and constant ups and downs in our lives, I she has been a patient mentor. I would also like to thank my committee members Dan Cosley and Lee Humphreys for constantly being available for their thoughtful and intellectual feedback. Having them has helped me overcome multiple challenges and together, I could not have asked for a better committee.

Many thanks also to professors and senior researchers who have helped shape this research. In particular, I would like to thank Sara Kiesler, Gilly Leshed, Gregorio Convertino, Helena Mentis and Aruna Balakrishnan. I would like to thank all the graduate students in Susan Fussell's group for thought provoking insights during my lab presentations. I especially thank Leslie Setlock for helping manage assistance and helping improve my quality of presentations. This work would not have been possible without the help of multiple research assistants who gave me an opportunity to mentor them, and worked with me to create multiple systems that could be deployed for rigorous lab experiments, in no particular order: Sarah Alabdullatif (Research Associate, KACST), Poornima Prabhu (Amazon), Si Chang (Salesforce), Cristina Guzman (Palantir), Kailin Koch (BSG), Samantha Reig (Cornell), Ruoyan Qin (Cornell), Steven Lam (Research Masters, Cornell), Karthic Ashokan (Oracle), Wit Tatiyanupanwong (Yahoo), Yiming Li (Epic), Eric Swidler (Athena), Wei Yuan (Amazon), Yifan Tong (IBM), Luke Goodman (GE), and David Marchena (Cornell)

This dissertation would not have been possible without the moral support of the ACM SIGCHI community, which is now an extended family. Unfortunately, mentioning every name would extend this acknowledgement beyond the dissertation length. In particular, I would mention my recent collaborators, Pawel Wozniak, and Erin Brady, who has been my source of welcoming hugs at CHI. Special thanks to all the faculty members and PhD students in the Department of Information Science at Cornell University for the conducive environment towards teaching and research.

Tremendous thanks to my family for their patience while I pursued research and who I have been unable to stay constantly in touch with for multiple years owing to the distance. I thank them for their affection and love. I would also thank Yahoo!, Bloomberg, and Google for sheltering me over the summers and exposing me to a diversity of projects. This material is based upon work supported by the National Science Foundation under Grant No. IIS-0968450. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

CHAPTER 1

INTRODUCTION

In April 2013, the annual Boston Marathon was rocked by two bomb blasts that killed three people and injuring several hundred of them (New York Times, 2013). As teams of police, government officials, federal agencies, and others worked to solve the crime, multiple challenges were discovered that have motivated this work. First, even though the prime suspects, two brothers, had already been on the terror suspect list for 18 months and U.S. intelligence analysts had noted their radicalization trips, these analysts failed to share the notes with their counterparts abroad, leading to a loss of important information. Second, redditors on the subforum Reddit Bureau of Investigation tried to solve the case by sifting through multiple official and unofficial streams of data. Once a redditor implicated someone, other redditors started looking for evidence to support the hypothesis and ignoring evidence that did not support it, a phenomenon known as confirmation bias. Redditors also publically implicated the wrong person, revealing that people, even large groups of people, cannot always perform a complex sensemaking task well. Finally, there was no joint information center to manage interrupting requests by 24/7 news cycle, and this led to confusion about what each agency was doing.

The Boston Marathon bombing is only one of many examples that highlight the challenges people face when they collaborate to analyze and synthesize complex data to predict or solve crimes, an area I will call *crime analysis* in the remainder of this dissertation. Accounts by retired intelligence analysts (Heuer, 1999) and more examples presented across the chapters of this dissertation show that problems such as a lack of information sharing between agencies (e.g., Pickton Report, 2007; Egger, 2002; Mentis et al., 2009), cognitive biases that lead analysts

to focus on the wrong information (e.g., Police Chief Magazine, 2009; Tversky & Kahneman, 1973; Nickerson, 1998 ; Xu & Chen, 2005), an overload of information (e.g., Simon, 1957), interruptions (e.g., Czerwinski et al., 2000; Latorella, 1998; Speier et al., 1999), and insufficient workforce (e.g., New York Times, 2013) make solving crimes and crime analysis difficult.

A number of computer-based tools have been developed to assist people with crime analysis (e.g., Gottlieb et al., 1994; Santos, 2016; Bier et al., 2010; Isenberg et al., 2012; Stasko et al., 2008; Kang et al., 2014; Convertino et al., 2009; Analyst's Notebook/i2, 2011; CoMotion/General Dynamics, 2011; Palantir, 2011; Nobarany et al., 2012). These tools can help support analysis, for example by classifying relevant information into categories (Bier et al., 2010; Gottlieb et al., 1994), tracking information flow (Analyst's Notebook/i2, 2011), looking for red flags in large volumes of data (Palantir, 2011), supporting social information sharing (Nobarany et al., 2012; Kang et al., 2014; Convertino et al., 2009), and creating visualizations of the data (Stasko et al., 2008).

Despite the promise of tools to support crime analysis, research is mixed as to their value. Laboratory studies have shown benefits to having these tools, for example sharing visualizations of analysis can improve the odds of solving a crime (e.g., Balakrishnan et al., 2008, 2010), and enabling discussion of information that disconfirms initial hypotheses can lead people to consider a wider pool of suspects (e.g., Convertino et al., 2008, 2011). However, in real world settings analysts are reluctant to use analysis tools because they find them to be overly complex (Chin Jr., 2009; Heuer, 1999) or fail to integrate well with their personal analytical process (Egger, 2002). Instead, analysts rely on simple tools like Google search (Cowley et al., 2005),

post-it notes and spreadsheets, as they lay them out freely onto flat surfaces to manipulate and categorize them (Chin Jr., 2009; Heuer, 1999)

The overall goal of this dissertation is to design and evaluate tools to support crime analysis that are both easy to use, such that analysts want to adopt them, and useful, in that they help overcome the cognitive and social biases that negatively impact analysis outcomes, as was the case in Boston Marathon Bombing. I present three tool prototypes that I designed, developed and deployed to help improve collaborative sensemaking.

Specific Research Goals and Approach

I take a human-centered design approach to the question of how best to support collaborative crime analysis. My designs are informed by previous literature on sensemaking (e.g., Pirolli & Card, 1993; 1995; Russell et al., 1993; Waltz, 2003; Weick et al., 1993) and the analytic process (e.g., Chin et al., 2009; Convertino et al., Heuer, 1999; Johnston, 2005) as well as earlier designs for analysis tools (e.g., Analyst's Notebook/i2, 2011; CoMotion/General Dynamics, 2011; Palantir, 2011, Stasko et al., 2008). I then try to tease out how different features of collaborative analysis tools could support sensemaking and information sharing while helping to reduce cognitive biases.

Specifically, I address three sets of research questions:

1. What are the benefits and costs of common features of analytic support tools (e.g., notepads, visualization tools, chat boxes, etc.)? Is it possible to streamline the design of these systems by identifying the most important features?
2. Can analytical tools improve information sharing across geographically distributed teams (e.g., notes, insights, hypothesis, evidence etc.)? Would automatizing

information sharing, such that analysts don't need to remember to explicitly do so, improve collaborative crime analysis?

3. Could an analytical tool help analysts to be more self-reflective about their group-level sensemaking processes (e.g., highlighting if we are attending to all potential solutions equally, etc.)? Would this increased self-reflection benefit the analysis process or its outcomes?

I use a mixed method approach to address these research questions. First, I iteratively design and evaluate a series of analytical tools, each of which incorporates features specifically designed to address particular problems analysts face such as information overload or cognitive tunneling. These tools incorporate a variety of design ideas, including modularity for customization, implicit information sharing, and sensemaking translucence. Second, I conduct a series of carefully controlled laboratory experiments to isolate the impact of each design decision on the sensemaking task performance and user experience. I collect both qualitative and quantitative data to provide a deep understand how people use the tools, including task performance measures, subjective survey data, user log data, qualitative feedback and interviews.

Research Contributions

This dissertation contributes to the fields of human-computer interaction (HCI), design, and computer-supported cooperative work (CSCW) by developing and testing new strategies for supporting collaborative analysis, though modular design, including implicit sharing, and sensemaking translucence. Specifically, this dissertation's contributions are:

1. The creation of the SAVANT analysis platform, which allows investigators to selectively turn individual features such as notepads and visualization on and off,

thereby allowing for tests of how these individual features influence the sensemaking process.

2. Evidence for the value of specific features of analysis tools, such as implicit information sharing and sense-making translucence, for successful collaborative problem solving.
3. A set of design implications for future collaborative analysis tools, including suggested design of a series of prototype analysis tools, can benefit collaborative analysis.

Outline

In the chapters that follow, I describe three laboratory studies that explore different challenges faced during collaborative sensemaking, and discuss my findings and their implications. In Chapter 2, I discuss related literature on sensemaking, collaborative sensemaking, design of tools for collaborative sensemaking, the associated challenges and solutions I pursued to overcome these challenges. Chapter 3 describes the design and laboratory evaluation of SAVANT – the prototypical tool I developed for solo sensemaking, to identify the value of customization and modularity of sensemaking features. Chapter 4 discusses how SAVANT was appropriated for collaborative sensemaking, and evaluated for implicit sharing of intermediate data analytic artifacts such as notes. Chapter 5 presents the design and evaluation of REFLECTIVA, a tool that leverages intermediate data analytic artifacts to make sensemaking more translucent. Finally, Chapter 6 summarizes my dissertation research, discusses limitations and provides implications for future work.

CHAPTER 2

BACKGROUND

This chapter describes previous work in areas that are pertinent to the design space of collaborative sensemaking. The chapter begins by discussing the general sensemaking process, the biases that can arise during sensemaking, and collaborative sensemaking. Then, I review the goals and features of existing sensemaking tools, with a focus on tools aimed to support collaborative analysis. Finally, I describe the motivation behind my design strategy and the features that are incorporated in systems presented in later sections of this dissertation.

The Sensemaking Process

Sensemaking, as defined by Russell et al. (1993), is the process of searching for a representation and encoding data in that representation to answer task-specific questions. The process of sensemaking involves multiple stages, including: extracting data, creating categories, and iteratively categorizing and encoding data into the categories in order to generate representations that can best describe what is known.

Pirolli and Card (1993; 1995) propose a model of sensemaking comprised of two loops: *a foraging loop* and a *sensemaking*. During the *information foraging phase*, analysts identify relevant documents and search these documents for clues and relationships. Interviews of professional analysts suggest that initial organization of documents and information is a major challenge yet essential for successful results (Johnston, 2005). In a crime scenario, a homicide detective must identify which documents are most relevant to solving the crime, pour through these to uncover key people, places, weapons, and motives, and uncover relationships among these entities. An important aspect of information foraging is that analysts will continue to

examine a source of information so long as it provides relevant information for their tasks; when that relevance declines, they will move on to the next information source.

During the *sensemaking phase*, analysts construct mental models of their findings (Pirolli and Card, 2005), pulling their observations and the patterns they uncovered together into a coherent narrative, which then iteratively guides future information foraging. Analysts can consider multiple schemas simultaneously, as alternative explanations for observed facts and relationships. For example, they might simultaneously consider hypotheses in which one person committed the crime for financial gain, and another committed the crime because of jealousy. They then keep both hypotheses in mind as they return to information foraging.

Finally, during the *decision-making phase*, analysts will typically choose one of their information schemas to act on. For example, homicide detectives might decide that the preponderance of evidence points to a single culprit with a single motive and an opportunity to commit the crime. They might then recommend that he be arrested and charged. However, such a process is fraught with multiple challenges for analysts and for designers alike. These challenges are further complicated by the patchwork of technologies that exist today and would exist in future, because we need to design for today with an eye towards the future. In the next sections, I will lay out the challenges in sensemaking, process of collaborative sensemaking, and how technological designs have impacted the collaborative sensemaking process.

Errors and bias in sensemaking

A key challenge for sensemaking in crime, intelligence, and many other domains is the vast amount of pertinent information. *Information overload* and a scarcity of time can make it difficult for analysts to thoroughly vet all the available information. Analysts often need to

satisfice (Simon, 1957), picking a good but not necessarily optimal path through the vast amount of available information. Although satisficing can improve efficiency, it can also lead to biases in the analysis. After noticing a few suspicious aspects of a suspect's report of his whereabouts, for example, an analyst might focus on confirming that the suspect is guilty rather than thoroughly considering alternative suspects. Susceptibility to *cognitive tunneling* or *confirmation bias* (Nickerson, 19980; Xu & Chen, 2005) is common in human thought processes (Pirolli & Card, 2005) and works against successful crime solving.

Analysts' sensemaking may also be biased by the availability of information schemas from their daily lives or previous analytic activities. This *availability heuristic* (Tversky & Kahneman, 1973) leads them to pursue certain paths through the data based on expectations that are not grounded in the data but in what they can easily recall from memory. For example, an analyst who has solved crimes in which a jealous partner or spouse was the culprit may decide to focus on a current victim's partner or spouse regardless of whether or not the evidence points in that direction. To counteract these biases, analysts have proposed techniques to push analysts to consider all the data. For example, Heuer (1999) proposed that intelligence analysts can overcome bias by using what he calls the "analysis of competing hypotheses" (ACH) approach, in which they generate multiple competing hypotheses and search for evidence for and against each one. In an observational study of intelligence analysts, Chin, Kuchar, and Wolf (2009) found that many reported using a process that involved comparing hypotheses. Another promising direction is techniques that assist in working with large and complex datasets, for example, by visualizing the analytical reasoning process (Cook & Thomas, 2005).

Collaborative sensemaking

Research in organization science by Weick et al. (1993) points out that sensemaking is not always a solo activity; instead, groups of people may socially co-construct views of the data. Furthermore, real world crimes, such as the Boston Marathon bombing or the 9/11 terrorist attacks (National Commission on Terrorist Attacks, 2004), are tackled by large teams of investigators who would ideally work productively together. This collaboration can be asynchronous, such as when one agency passes documents along to another, or synchronous, such as when different agencies work together in real time to solve a case. Collaborating in the sensemaking processes can be advantageous since partners can leverage each other's cognition and insights to solve hard problems (Goyal & Fussell, 2015; Hayne et al., 2011; Willett et al., 2011). Multiple analysts may have different access to documents, and with more readers there is a greater ability to sift through large amounts of data and identify patterns.

Like individual sensemaking, collaborative sensemaking requires information foraging, information schematization, decision-making, communication of insights and facts, and collaborating to identify a solution (Chin et al., 2009). But collaborative sensemaking can be especially challenging because analysts need to coordinate their tasks and workflows, share information with one other, and perform joint decision-making (Waltz, 2003). For example, analysts are often reluctant to exchange information and insights for fear they might be wrong (Heuer, 1999). In addition, the exchange of incorrect information can lead to poorer outcomes due to what Kang and Kiesler have termed *teammate inaccuracy blindness* (Kang et al., 2014). That is, analysts treat all information from a partner as valid and useful, regardless of its actual

quality. As a result, if one member of a team of analysts prematurely focuses in on an incorrect suspect, the other is likely to follow.

Even when collaborators do share information with one another, they may have difficulty establishing common ground or mutual knowledge of what that information means (Clark & Brennan, 1991). There are multiple reasons why analysts might not be able to reach common ground (Klein et. al., 2005; Laurence et. al, 2016), requiring continuous repair, and interpretation.

Problems of information overload may be exacerbated when analysis is collaborative rather than solo, because new information from partners may come at unpredictable moments. In general, prior work has shown that random interruptions can have significantly negative impact on task completion time (Czerwinski et al., 2000), task performance (Latorella, 1998), decision-making (Speier et al., 1999), and affective state (Bailey et al., 2006; Bailey et al., 2006; Zijlstra et al., 1999).

Sensemaking for crime analysis

As discussed previously, the intelligence analysts iterate multiple times through *foraging*, *sensemaking*, and *decision-making phases*. However, this sensemaking process in crime analysis is a specific kind of sensemaking. The analysts must inductively parse documents to discover the breadth of potential suspect choices and then use the clues uncovered in the discovery process to deduce the right criminal of all the suspects. This combination of induction and deduction sets crime analysis as a special sensemaking task. Other sensemaking tasks, like medical sensemaking (Mentis et. al, 2009, Paul et. al, 2010), may focus on induction or deduction conversely. This careful iterative dance between induction and deduction makes collaborative

sensemaking even harder because, while one partner might be still be inductively pursuing suspect options, the other partner might have already deduced the criminal. How may then the two analysts collaborate and negotiate the correct criminal when one has not yet fully discovered the solution space?

Summary

As shown in the sections above, sensemaking is an iterative process in which individual analysts or teams of analysts iteratively sift through data and reason about that data. It can be negatively impacted by cognitive biases that lead analysts to focus on some pieces of evidence and ignore others and by social processes that make analysts reluctant to share information that their teammates need. Further, it is evident that information and cognitive overload created due to paucity of time and resources present the challenge of solvability in *crime analysis*. This refers to how it remains challenging to solve crime cases due to In the next section, I discuss analysis tools that have been developed to address these problems

Tools for Collaborative Sensemaking

A wide variety of tools have been developed to support the sensemaking process and to try to reduce the negative effects of cognitive biases, information overload, and social dynamics, including both research prototypes (e.g., Chin Jr. et al., 2009; Convertino et al., 2008; Shrinivasan et al., 2008; Stasko et al., 2008) and commercial analysis products (e.g., Analyst's Notebook/i2, 2011; CoMotion/General Dynamics, 2011; Palantir, 2011). For example, tools can help analysts visualize and manipulate data at different levels of granularity to detect links between objects in large datasets and construct alternative hypotheses (Chin Jr. et al., 2009; Convertino et al., 2008; Shrinivasan et al., 2008; Stasko et al., 2008), as well as collect and

arrange data and notes for later reference (Andrews et al., 2010; Pioch et al., 2006; Wright et al., 2006).

Many existing analysis tools have taken the approach of incorporating features to address a broad array of analyst needs. For example, Bier et al.'s (2010) Entity Workspace system provides a reading area, search tools, collection overviews, and an evidence notebook to support the analysis of large document collections. More sophisticated features within the evidence notebook allow analysts to organize entities, create information structures that can be collapsed and expanded, visualize evidence that emphasizes events and documents. There is also a notification system that finds and displays entities of mutual interest to multiple analysts.

In the remainder of this section, I review some of the common features found in analysis tools.

Visualizing data and relationships among entities

Tools for visualizing and detecting links between objects in large datasets have been specifically suggested to assist and improve analysts' work. Such visualizations can help users aggregate and abstract activities (Heer and Agrawala, 2008), provide a view into the dataset using a shared network diagram (Balakrishnan et al., 2008), show a timeline (Ganoe et al., 2003), or present a user activities list (Heer et al., 2009).

Stasko et al.'s (2008) Jigsaw tool, for example, is designed to help the sensemaking process through flexible visualization and manipulation of the data. A special "shoebox" area can be used to collect pertinent data for later reference. Shrinivasan et al.'s (2008) Aruvi integrates a suite of visualization tools with access to the data at different levels of granularity. CrimeNet Explorer (Xu et al., 2005) creates and visualizes concept maps based on word co-occurrence

between crime report documents, with the goal of helping crime analysts detect members of criminal networks such as gangs. CACHE (Convertino et. al., 2008, 2011) explicitly incorporates Heuer's Analysis of Competing Hypotheses (ACH) technique (Billman et al., 2005; Heer et al., 2008), allowing analysts to build and evaluate matrices linking evidence to alternative hypotheses.

Other visualization tools aim to help collaborators achieve common ground using collaborative visualizations (Chung et al, 2010; Heer and Agrawala, 2008; Isenberg et al., 2012; Janssen et al., 2007; Stasko et al., 2008; Tversky et al., 1975). Chuah and Roth (2003)'s Command Post of the Future, for example, tries to establish common ground between analysts through a combination of explicitly shared objects and events, representations of level of attention directed to objects, depiction of goals for analyzing objects and events, representation of interpretations and thoughts through annotations and sketches, and representation of object history.

Structuring and manipulating information

Many of these tools also help analysts structure and manipulate information as well as prepare reports on their findings. Sandbox (Wright et al., 2006) allows analysts to flexibly arrange notes and other information as they form information schemas. POLESTAR (Pioch & Everett, 2006) allows analysts to preserve snippets of textual data in a central archive, use them to construct arguments, and link to them in final analysis reports.

Several tools allow analysts to structure the data collaboratively. For instance, Bier et al. (2010) and Hayne et al. (2011) employ shared workspaces where analysts may classify and categorize existing information for future use into distinct categories. Co-located tools for

analysis have also been developed to help users identify pertinent information by searching, sorting, filtering and creating schemas (Isenberg et al., 2012); or find relevant pieces of information to generate story lines in location aware cross-device manipulation setups (Wozniak & Goyal et. al, 2016).

Communication tools

Leveraging partners' insights requires sharing of insights and subsequent awareness of these insights. Prior research in sensemaking tools has shown that explicit communication through chats or comment threads (Heer et al., 2009) or annotations (Kang et al., 2012) helped analysts build a shared mental model during collaborative analysis.

Other tools, meant for more general tasks like group brainstorming can also be helpful for collaborative analysis. For example, GroupMind (Shih et al., 2009) balances between formal and informal models of the data through a process of incremental formalization; Livenotes (Kam et al., 2005) suggests collaborative note-taking for improved understanding, and TEAM STORM (Hailpern et al., 2007) allows analysts to work on multiple ideas in parallel to facilitate re-interpretation and reflection on the action.

Shared workspaces

Shared workspaces have been shown to improve shared understanding and awareness in a wide variety of domains (Gergle et al., 2004; Gutwin & Greenberg, 1998) by promoting exchange of information and data with others (Hayne et al., 2011), improving common ground (Willett et al., 2011) and increasing awareness of the status of the analysis task and others' activities in the task (Convertino et al., 2009; Pioch & Everett, 2006). Shared workspaces have primarily been researched from an explicit sharing perspective where analysts consciously

choose to share their insights at the appropriate point in the collaborative sensemaking cycle (Chuah & Roth, 2003; Convertino et al., 2008; Nobarany et al., 2012), although as noted earlier, analysts may be reluctant to share their ideas and information before until they are confident in their insights.

Other tools have been incorporated into shared workspaces for synchronous collaborative sensemaking space. For example Andrews et al. (2010) and Vogt et al. (2011) examined the effects of large displays on the sensemaking process. Wong et al.'s IN-SPIRE and HI-Space (2006) use a collaborative mobile, tabletop and wall display tool for dynamic analysis in group environments using visual analytics for collecting evidence and considering parallel competing hypotheses. Hayne et al. (2011) uses Design Cards across shared whiteboards for cognitive alignment between team members

Notifications, reminders and recommendations

Supporting collaborative analysis could include reminding analysts to view their partners' analysis, as in AnalyticStream (Nobarany et al., 2012) or recommending relevant pieces of information from their partner (Bier et al., 2010). However, biases owing to personal beliefs may inhibit taking advantage of such features (e.g., Constant et al., 1994; Jarvenpaa et al., 2000; Lee et al., 2007) or lead to groupthink and/or cognitive tunneling (Willett et al., 2011).

The tools designed to enable real-time notification vary from notification of each activity by the partner (Brush et al., 2002) to shared workspaces (Gergle et al., 2004; Gutwin & Greenberg, 1998) where others' activities in the task are listed out (Convertino et al., 2009; Pioch et al., 2006). Instead of explicitly notifying activities, tools may also remind analysts to learn about their partners' analysis (Nobarany et al., 2012).

Minimizing the impact of interruptions

As noted earlier, random interruptions can have significantly negative impact on task completion time (Czerwinski et al., 2000), task performance (Latorella, 1998), decision-making (Speier et al., 1999), and affective state (Bailey et al., 2006; Bailey et al., 2006; Zijlstra et al., 1999). Consequently, researchers have pursued multiple ways to overcome these negative impacts. One strategy is to schedule interruptions at breakpoints between tasks (Bailey et al., 2006; Bailey et al., 2008), which are identified by modeling the tasks. Alternatively, one may create statistical models that detect such breakpoints based on user activity (Bailey et al., 2008; Fogarty et al., 2005; Horvitz et al., 1999). Finally, researchers have also attempted to identify the scheduling using correlation between interruptible moments and physiological data, such as pupil-dilation (Bailey et al., 2008), or studying how multiple sensor inputs can be used together to interrupt non-relevant distractions in a software engineering tasks (Züger et al., 2015).

Perhaps one of the more recent tools aimed at interruption scheduling is OASIS, a system that defers notifications until breakpoints are reached to reduce interruption costs (Bailey et al., 2008). The tool identifies the level of breakpoint needed for a notification, and then accordingly interrupts when a notification arrives. While this tool has been shown to work with six users on non-sensemaking tasks, it requires continual user-activity monitoring, via software installation at workplace, which has privacy implications. Further, the tool also required extensive training thought automated and manual coding.

Summary

In summary, a variety of analysis tools have been designed to date. But while the inventors of the tools report that they are beneficial for analysis, the majority of them have been

evaluated only informally, with a handful of users (e.g., Bier et al., 2010). Further, many tool evaluations were conducted with all features available simultaneously. For example, Kang et al. (2011) conducted a controlled study with their Jigsaw tool to understand the influence of its components on the analytical process. Design implications derived from their study include facilitating clue-finding, supporting evidence marshalling, and allowing flexible note taking.

While such studies are valuable, they do not shed light on the unique benefit or cost of any one single design feature to the analytical process, nor do they in most cases provide sufficient data for statistical testing of the effects of a given feature. Furthermore, analysts often fail to adopt these tools. Cowley et al.'s (2005) study of intelligence analysts' tool use showed that they spent the preponderance of their time either using a web browser to search for information or a word processing program to compile the results. They seem to have done most of their hypothesis generation and analysis in their heads, leaving no documentation trail. Similarly, Chin et al. (2009) found that analysts preferred either paper graphs and diagrams or PowerPoint and Excel for their analyses over specialized analysis support tools.

This Dissertation: Designing Technology for Sensemaking in Crime-Solving

In the first two sections of this chapter, I showed that analysis is a complicated sensemaking process that is made even more complicated when analysts collaborate, and that a wide variety of tools have been created to help address these sensemaking challenges, often by including large sets of features. At the same time, analysts still prefer simple tools like web search and Excel documents. In this section, I briefly describe my overall design goals and the specific strategies I used in the research prototypes presented in later chapters of this dissertation.

My work takes a human-in-the loop approach to designing tools to improve collaborative sense-making. In other words, my tools aim to improve how well analysts can perform sensemaking and how well they collaborate with one another, rather than, say, replacing human efforts with deep learning or other computational approaches. In particular, I focus on iteratively designing tools to address three specific challenges:

- C1. Design a sensemaking tool that allows researchers to evaluate the benefits and costs of specific features in addition to its usefulness as a whole.
- C2. Design a sensemaking tool that facilitates information sharing by helping analysts overcome social barriers to sharing
- C3. Design a sensemaking tool that enables the analysts to be more reflective and aware about their sensemaking process and avoid cognitive fixation.

I briefly discuss each of these challenges and my approach to addressing them below.

Evaluating features of sensemaking tools

As noted earlier, many analysts are reluctant to use current analysis tools because they find them overly complex (e.g., Cowley et al., 2005; Chin et al., 2009). However, it is difficult to understand how to simplify these tools because most have been designed to have many features and tested with small number of users employing all the features of the system simultaneously (e.g., Bier et al., 2010; Kang et al., 2011). My first goal was thus to create a modular platform that enabled testing and evaluating specific features individually. My system SAVANT, presented in Chapter 3, combines concepts from previous visualization tools (e.g., Billman et al., 2005; Heer et al., 2008), enabling the organization and visualization of data at different levels, and visualizing links between objects that otherwise couldn't be detected. I

show how this platform can be used to compare the value of two different sensemaking features, a shared note space and a visualization tool

Designing to facilitate information sharing

As noted earlier, analysts can be reluctant to share intermediate products of their analysis due to concerns about being correct, but partners may benefit from receiving this information earlier in the sensemaking cycle (Egger, 2002 ; National Commission on Terrorist Attacks, 2004; Cabrera, 2002). In addition, the need to decide when and to whom information should be shared can create additional cognitive load on analysts. To respond to this challenge, I explored whether *implicit sharing*, in which the system automatically shares insights between analysts, can better support collaborative analysis. In Chapter 4, I will discuss a collaborative version of SAVANT that was designed for evaluating the value of implicit sharing in addition to explicit sharing. I found that both explicit and implicit sharing are complimentary channels that benefit from each other's presence but either one is not sufficient towards identifying the hidden pattern.

Designing for sensemaking translucence

Prior work shows that even when analysts freely exchange their ideas, biases such as cognitive tunneling can lead to poor quality decisions (Police Chief Magazine, 2009; Mentis et al., 2009; Willett et al., 2011; Kang et al., 2014). To facilitate the exchange of insights while simultaneously discouraging cognitive tunneling, I developed what I call a *sensemaking translucence* interface in my REFLECTIVA tool presented in Chapter 5. This interface consists of two main features: a *hypothesis window* and a *suspect visualization*.

The *hypothesis window* is similar to the Alternative Competing hypothesis (ACH) interface (Convertino et al., 2008), in which users explicitly share their hypotheses and evidence

so that they can maintain awareness of one another's insights and develop a joint mental model of the case. The hypothesis window is also designed to help reduce confirmation bias by including fields for reporting evidence that disconfirms each hypothesis (Goyal et al., 2016).

The *suspect visualization* depicts the joint attention paid to each suspect thus far in the analysis and encourages collaborators to distribute their attention across multiple suspects instead of focusing prematurely on a single suspect who might not be the actual culprit. The suspect visualization changes automatically as analysts mention suspects in their hypotheses, notes or chat conversations.

Summary

This chapter has reviewed behavioral science work showing the challenges of individual and collaborative sense-making, discussed human-computer interaction tools that have been developed to facilitate crime analysis and problems with assessing the value of their features, and outlined the design goals of my dissertation. The next chapter discussed the first contribution of this dissertation, the SAVANT sensemaking tool that allows researchers to investigate the benefits and costs of individual features such as visualizations, timelines, shared notepads, and so forth.

CHAPTER 3

SAVANT: A MODULAR TOOL TO SUPPORT CRIME ANALYSIS

As discussed in Chapter 2, when analyzing crimes, detectives and other police personnel examine witness and suspect interviews, crime scene reports, the coroner's findings, and many other documents in order to detect an underlying pattern and identify a culprit (Gottlieb et al., 1994; Santos, 2016). While existing analysis tools can improve sensemaking (Bier et al., 2010; Isenberg et al., 2012; Stasko et al., 2008; Kang et al., 2014; Convertino et al., 2009; Analyst's Notebook/i2, 2011); CoMotion/General Dynamics, 2011; Palantir 2011), analysts find these tools overly complex (Heuer, 1999; Johnston, 2005; Chin et al., 2009; Goyal et al., 2013). However, determining how to simplify analysis tools is difficult because most studies have tested systems as a whole set of features rather than evaluated individual components (Stasko et al., 2008; Bier et al., 2010; Convertino et al., 2009; Convertino et al., 2011). It is therefore not clear which components could be dropped from the system.

In this chapter, I first present SAVANT (Sensemaking and Analysis using Visualization and Note-taking Tool), a system designed to support the sensemaking process in crime analysis. SAVANT was designed to provide a variety of specific analysis features that could be turned on and off individually in order to test the value of each feature. Next, I use SAVANT to assess the value, separately and in combination, of two kinds of analysis features: a *visualization tool* that captures relationships among documents, and a *notepad tool* that allows users to take notes and link them to the original documents. Based on previous research about the utility of visualizing links between information objects and recording and summarizing information for the analysis process (Balakrishnan et al., 2008; Balakrishnan et al., 2010), I hypothesized:

H1: Analysts with a visualization tool that captures relationships among documents will be better able to solve a crime analysis task than analysts without a visualization.

H2: Analysts with a note-taking tool that allows them to mark down their observations and snippets of text from documents will be better able to solve a crime analysis task than analysts without a notepad.

While I expected each tool individually to be beneficial to analysis, it is less clear that having both together will improve performance over having either one alone. It might be that analysts will find manipulating two tools, in addition to finding and reading documents, too mentally taxing (e.g., Xie et al., 2000). Alternatively, the two tools might work synergistically with one another, leading to better performance than with either one alone. Thus, I posed this research question:

RQ: What will be the benefit of providing both a visualization and a notepad tool?

I examined the hypotheses and research question in a laboratory study in which students acted as detectives solving a serial killer crime. The visualization significantly improved participants' ability to detect key clues in the documents and identify the serial killer, whereas the notepad did not. Having both tools available provided no benefit over having just the visualization tool. The ways in which participants used the tool and their feedback helped me understand how different design features can interact with each other, and reduce overall task performance.

In the remainder of this chapter, I first describe the design of the SAVANT system, including the motivation for each of these features. Then, I present the design, method, and results of my laboratory experiment. I conclude with a discussion of the implications of this work for the design of analysis tools.

The Savant System Design

I designed SAVANT to aid sensemaking in analysis and overcome the difficulties of cognitive tunneling without sacrificing efficiency. I describe here the principles I followed in its design based on existing design principles for exploratory analysis (Perer et al., 2008) as well as technical details of the implementation.

SAVANT's user interface consists of five primary tools, each positioned in one pane, as shown in Figure 1. The directory (a) is a folder-like hierarchy of the document library, organized by cases. Double-clicking a document opens it in the reader pane (b), and multiple documents can be opened as reader pane tabs. The visualization (c) shows documents as nodes, with the currently open document in the center and the documents linked to it through word-co-occurrence processing. The visualization can be zoomed, panned, and nodes can be moved around. The user can also open a document from the visualization by double-clicking a node. The notepad (d) is a text editor in which text can be copied and pasted directly from an open document as well as freely typed in. Documents can be dragged from the directory to the favorites (e) to be bookmarked for later reference. Movable dividers allow resizing all of these parts.

Beyond additional assisting tools, I implemented this principle by devoting a substantial part of the screen estate to reading documents (in Figure 1, I moved the dividers to magnify the visualization and notepad panes to show details), with scrolling capabilities as well as the ability to adjust the size of the reader pane for large documents. Quickly switching between multiple open documents is available through a tabbed layout of the reader pane. The directory also shows which documents are currently open.

Another important goal of the design is to aid sensemaking when analyzing a large set of data, by *allowing the extraction of key information*, thereby reducing cognitive load. The user can highlight important text they find while reading a document in the reader pane, which is automatically copied to the notepad for easier access in the future instead of having to search through the entire document set again. In addition, the notepad serves for jotting down comments and hypotheses. In early prototypes, notepad comments were structured to contain links to the document from which they were extracted, time stamp, and other information. However, pilot testing demonstrated that unstructured free text, similar to scratch paper, was preferred. Another way to pull out important information at a larger granularity is by dragging an entire document from the directory to the favorites pane. Documents in the favorites serve as bookmarks and can be visually rearranged within the favorites pane.

Perhaps the key design principle for helping analysts make sense of and solve a complicated problem is to help them *discover relationships between documents*. The design of the visualization is based on co-occurrences of words between any two documents. For any single document, the visualization shows all the documents in the dataset that contain words in common with the active document as edges between the document nodes. The thickness of an

edge is based on the number of unique words in common between the joint documents, using TF-IDF. Hovering the mouse pointer over an edge reveals a tooltip with the words that generated the edge. The left of the visualization presents a list of words that the active document has in common with the other documents in the dataset. Selecting one or more words in the list serves as a filter, showing only the edges between documents that share the selected words, as shown in Figure 2. The visualization was implemented using the open source Radial-Tree component of ProtoVis (Heuer, 1999), which integrates using XML with the rest of the Java code.

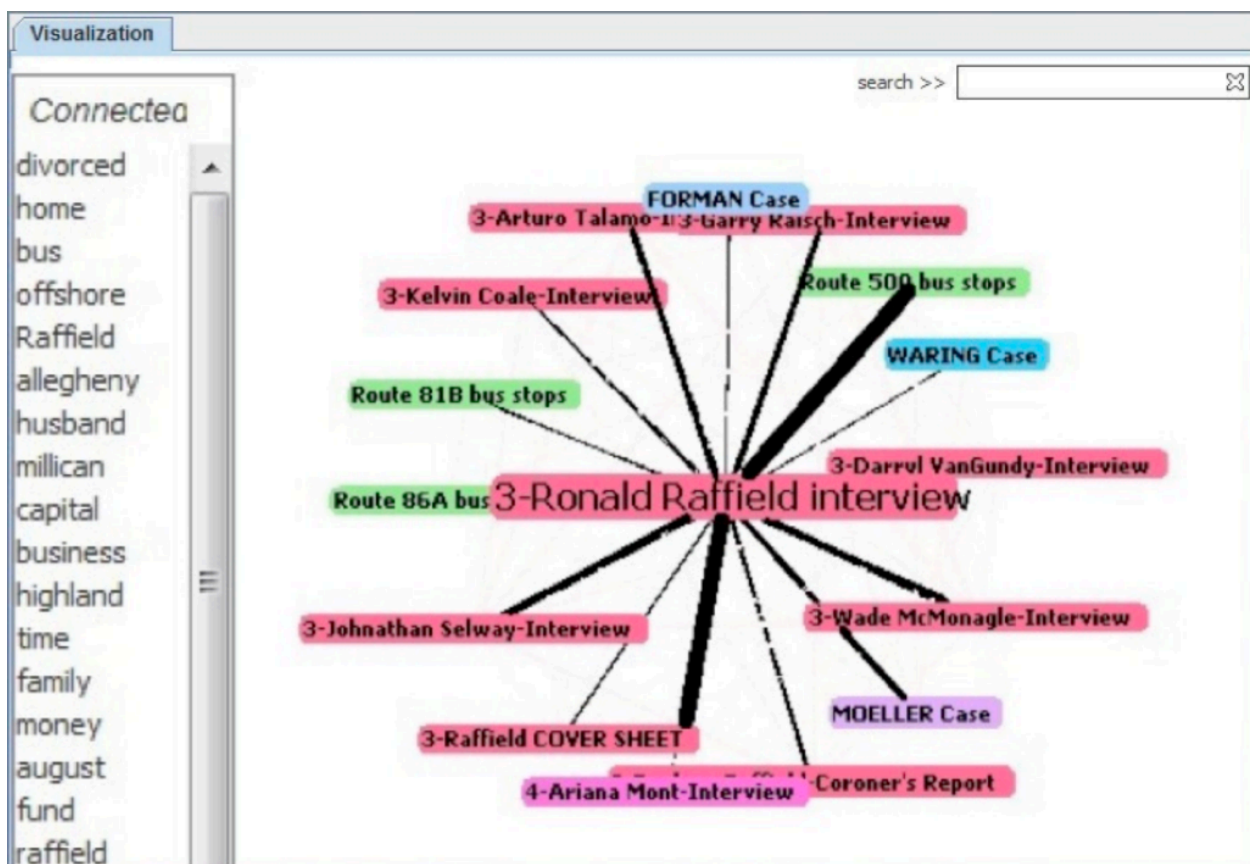


Figure 2. The visualization showing filtered results with documents from distinct cases that share a specific word.

In many analysis tasks, especially in crime investigation, it is seeing the links between documents from separate categories that is necessary to solve the task. For example, noticing that two separate crimes in different times and geographical areas share similar patterns in the crime scene findings may raise a flag to search for a common culprit. Documents in the visualization (and in the directory, for consistency) are color-coded based on the cases or categories to which they belong. This color-coding helps to reduce cognitive tunneling by highlighting documents from separate cases that are connected through edges in the visualization, as shown in Figure 2.

Finally, given a large number of documents, reviewing all of them may be impossible given time constraints. In order to *find important information*, another design principle I followed, I built a search feature in the directory and in the visualization. These are two separate search functions: in the directory, the search function highlights those documents that match the search query, whereas the visualization only shows pairs of documents that both match the searched query. To support the search functions as well as constructing the links between documents in the visualization, the text in the documents is indexed using the open-source LUCENE indexer with a stop-word list (Hatcher et al., 2004).

The SAVANT system was designed and developed iteratively, testing informally the value of each component and refining them accordingly. To evaluate, specifically, the visualization and the notepad tools, I carried out a laboratory experiment, described next.

Laboratory Study Comparing Visualization and Notepad Tools

To evaluate the features of SAVANT I asked participants to work on solving crime problems in which evidence for a serial killer were hidden among various documents. I manipulated the presence of two key features of the interface: the visualization tool, and the

notepad. This 2 by 2 between-subject design resulted in four conditions: visualization and notepad (VN); visualization, no notepad (V); no visualization, notepad (N), and; no visualization, no notepad (Control). The other components of the interface (directory, reader pane, and favorites) were available to all participants. Participants spent one hour working on the task, after which I measured their use of the various interface features, their cognitive load, and their ability to identify key clues and the serial killer.

Participants

Forty-one Cornell students (38 undergraduate and 3 graduate, 48% female) were recruited through flyers on campus. One participant was dropped from the analysis for lack of fluency in English. Only 29.3% reported watching crime dramas such as CSI or Law and Order once a week or more and only 14.6% read crime novels once a week or more. Participants were paid \$22.50 for the 90-minute study.

Materials

I adapted crime analysis training materials, crime case documents (Appendix A) and surveys (Appendix B) from Balakrishnan and colleagues (Balakrishnan et al., 2008; Balakrishnan et al., 2010).

Training materials. Training materials included a one-page description on how to solve crimes (e.g., look for motive, opportunity, and lack of alibi), a practice task involving a laptop theft, and a worksheet for reporting the possible culprit in the practice task.

Task documents. Case documents consisted of two “active homicide” case folders, six “cold cases” (unresolved cases from the past), and general information. Each active case folder included six documents: a cover sheet, four witness and suspect interviews and a Coroner’s

report. Each cold case included a one-page document that summarized details about the victim, time, method, and witness interviews. Four of the six cold cases were “serial killer” cases, such that they demonstrated similar crime patterns (e.g., killed by a blunt instrument).

However, the key clue to naming the serial killer was hidden in one of the active homicide interview documents: the culprit’s alibi for the active homicide was being observed riding the bus and carrying a toolbox. Additional documents included maps of the city, bus route diagrams, crime statistics, and a police department organization chart. Consulting the maps and bus routes enabled seeing that all cold cases were located on the bus route ridden by the serial killer. All active and cold case documents as well as the bus routes were included in the system’s directory pane; the maps and the organization chart were provided on paper for better readability.

Crime-solving tools. To help participants solve the crimes, participants were given paper copies of an MO (*modus operandi*) worksheet in which they could indicate key details pertaining to the crime such a motive, opportunity, and alibi, a timeline, and a suspect list where they could enter names of suspects.

Post-task report. A paper-based post-task report included spaces to indicate for each crime the prime suspect (if identified), any known attributes of the suspect, the victim(s), the MO, and location.

Post-task survey. A modified version of Balakrishnan et al.’s (2010) online survey was distributed after participants completed the analysis task. The survey included demographic questions about gender, occupation, age, ethnicity, and education level. In addition, five seven-point scales adapted from the NASA TLX survey (Hart et al., 1988), were used for rating how much mental demand, temporal demand, effort and frustration participants felt during the task

and how well they thought they performed. The survey also included 22 multiple-choice questions about the crimes designed to measure how many key clues they had identified.

I also developed a series of questions about features of the interface. I used five-point scales to rate the usefulness of the various components of the system – the directory, reader pane, favorites, visualization and notepad, as well as the search feature. Participants typed in additional thoughts about the user interface and their experience using open-ended questions.

Procedure

Upon arriving at the lab, participants were randomly assigned to one of the four conditions, resulting in 10 participants per condition. They were first given an overview of the study, signed consent forms, and performed the practice task. Participants were then seated at a computer with a 25” monitor presenting the SAVANT system in full screen. The experimenter demonstrated each feature of the system available in the participant’s condition (directory, reader pane, and favorites for all participants; visualization and notepad where applicable). With the experimenter’s guidance, participants practiced interface tasks such as opening a document, highlighting text and copying it into the notepad, searching the directory and clicking on filters in the visualization to ensure they understood how each feature worked.

Participants then proceeded to work on the homicide cases. Unlike previous studies that used the same set of documents (Balakrishnan et al., 2008; Balakrishnan et al., 2010), participants were not told that a serial killer was hidden in the documents, only that they had one hour to solve as many cases as they could of the two active and six cold cases. While working on the task, the SAVANT system logged every time the mouse pointer entered and exited the different components (directory, reader pane, favorites, visualization, or notepad). After 60

minutes or earlier if the participant indicated that they had finished the task, they switched to completing the post-task report and the survey, after which they were paid and excused. The whole session lasted 90 minutes.

Measures

Performance. I assessed performance based on the post-task report. I scored whether or not participants identified and named the serial killer, indicating which crimes he was responsible for. Participants received a score of 1 if they correctly identified the serial killer and 0 otherwise.

Clue detection. As another measure of performance, I assessed participants' detection of key clues in the crime scenario. Based on the post-task report, I counted how many clues, out of 9, they described in the report (*clue recall*). In addition, from the post-task survey I counted the number of correct answers of 10 questions pertaining to the serial killer, each having one correct answer out of four (*clue recognition*).

Connections between documents. One question asked participants if they identified connections between the cold cases, and another question asked whether participants saw a connection between the cold cases and the active case holding the key clue. A participant's score of 1 indicated answering "yes" on both questions, 0.5 for answering "yes" on one question, and 0 for answering "no" on both.

Cognitive load. Four of the five NASA-TLX scales (excluding perceived performance) were highly correlated and formed a reliable scale (Cronbach's $\alpha = .80$); responses were averaged to create my measure of cognitive load.

Tool usefulness. Participants' ratings of the usefulness of the five key components of the interface (directory, reader pane, favorites, visualization, notepad) and the three search capabilities (directory, reader pane, visualization) on five-point scales were used as direct measures of their tool preferences.

Tool use. I approximated the time spent using each component available to participants in the tool by calculating from the system logs the total time duration that the mouse pointer visited that pane. I removed two values that were more than 2.5 SD above the mean, which could have been errors. I divided these times by the total amount of time a participant spent working on the task to create percentage time measures.

Results

Overall, the findings suggest that using the visualization tool helped solving the analysis task, supporting H1: participants using the visualization were more likely than those not using it to find the serial killer, they detected more clues, and they were better able to see relationships between documents. Conversely, participants using the notepad were less likely to solve the task, detected fewer clues, and were less able to see relationships between documents, rejecting H2. In response to the research question, I did not find evidence that the visualization and notepad tools together had an added benefit beyond the visualization tool alone. Still, my findings point to ways in which participants found each of these tools useful for the analysis task, and to ways in which these tools can be improved.

Task performance

Overall, participants with the visualization tool (80%) were more likely to succeed at identifying the serial killer than participants without the visualization tool (40%). In contrast,

participants with the notepad were somewhat less likely to identify the serial killer (50%) than participants without the notepad (70%), and participants who had both the notepad and the visualization tool were less likely to identify the serial killer (70%) than those with the visualization tool alone (90%) (See Figure 3). In other words, it seems that the visualization improves task performance, whereas the notepad undermines performance.

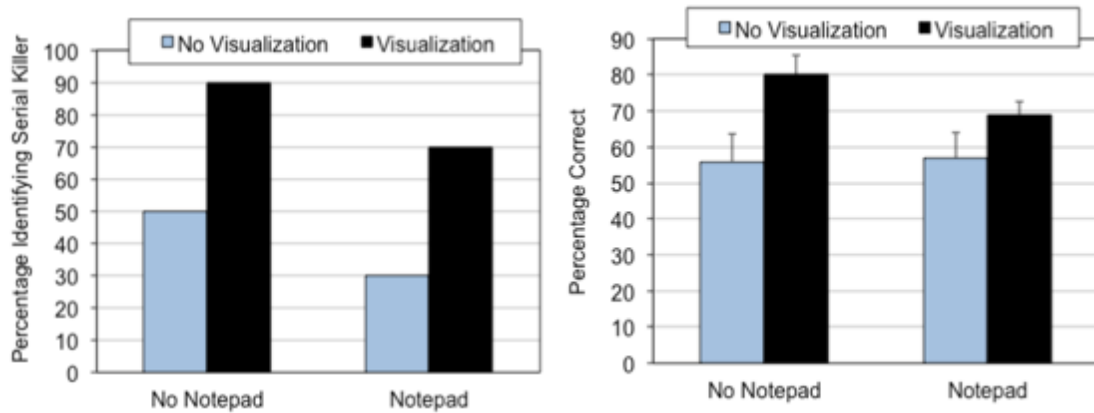


Figure 3. Percentage of participants correctly identifying the serial killer as a function of visualization and notepad condition.

I tested these relationships using a binary logistic regression model in which success was a binary dependent measure and visualization condition, notepad condition, and their interaction were used as predictors. Results showed a borderline significant effect of visualization condition in the direction shown in Figure 3 ($B=2.20$, $S.E.=1.23$, $Wald=3.20$, $p=.07$), but no effect of notepad ($B=-.85$, $S.E.=.94$, $Wald=.82$, $p=.37$) and no interaction ($B=-.50$, $S.E.=1.57$, $Wald=.10$, $p=.75$). These results provide partial support for H1, and no support for H2.

Clue detection

I measured clue detection in two ways: free recall of the nine key clues in the post-task reports and clue recognition in 10 multiple-choice questions in the post-task survey. These two measures were highly correlated with one another ($r=.70$, $p<.001$), and both were also highly correlated with successfully identifying the serial killer (for clue recall, $r=.77$, $p<.001$; for clue recognition, $r=.68$, $p<.001$). The findings on both clue recall and clue recognition support H1, do not support H2, and show no specific benefit or detriment for using the visualization and the notepad tools together.

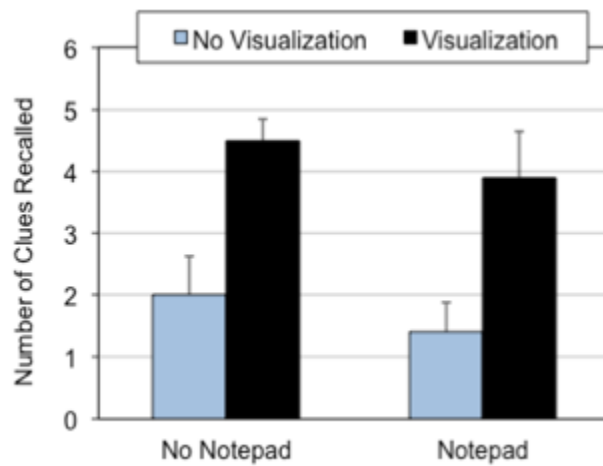


Figure 4. Number of clues recalled (out of 9) as a function of visualization and notepad condition.

(Error bars represent standard errors of the mean.)

Clue recall. Clue recall scores were analyzed in a 2 (visualization condition) by 2 (notepad condition) ANOVA. When participants had the visualization tool available, they recalled significantly more clues ($F[1,39]=19.43$, $p<.001$). Having the notepad tool available, however, had no significant effect ($F[1,39]=1.12$, $p=.30$). There was also no significant interaction between conditions ($F[1,39]<1$, n.s., as shown in Figure 4).

Clue recognition. The results using clue recognition in the multiple-choice questions showed a similar pattern. A 2 (visualization condition) by 2 (notepad condition) ANOVA showed that having the visualization tool led to significantly greater performance on these questions ($F[1,39]=8.49, p<.01$) but having the notepad present had no effect ($F[1,39]<1$, n.s.). There was also no significant interaction between visualization and notepad condition, as shown in Figure 4 ($F[1,39]<1$, n.s.).

Relationship detection

Similarly, having a visualization tool significantly increased participants' ability to detect relationships among documents, whereas having the notepad had no effect. A 2 (visualization condition) by 2 (notepad condition) ANOVA showed a significant effect for visualization tool ($F[1,39]=14.57, p=.001$), no effect of notepad ($F[1,39]=1.62, p>.20$) and no interaction ($F[1,39]=1.62, p>.20$). These findings, again, support H1, show no support for H2, and provide no evidence for benefit of both tools together over one alone (RQ).

Tool use

The results thus far refer to the explicit value of the visualization and notepad tools for task performance measures. Beyond task performance and to inform future design, however, I was also interested in understanding *how* these tools were used. One of these aspects is the time spent on each tool.

All participants had the option of working with three of the tool components: directory, reader pane, and favorites. Depending on their condition, participants also had the option of working with the visualization or notepad tool components. As can be seen in Figure 5, participants spent the most time reading the case documents, as I initially expected, purposely

designing the reader pane to be the largest element of the user interface. When the visualization tool was available they spent about 11% of their time on it, and when the notepad was available they spent about 9% of their time on it.

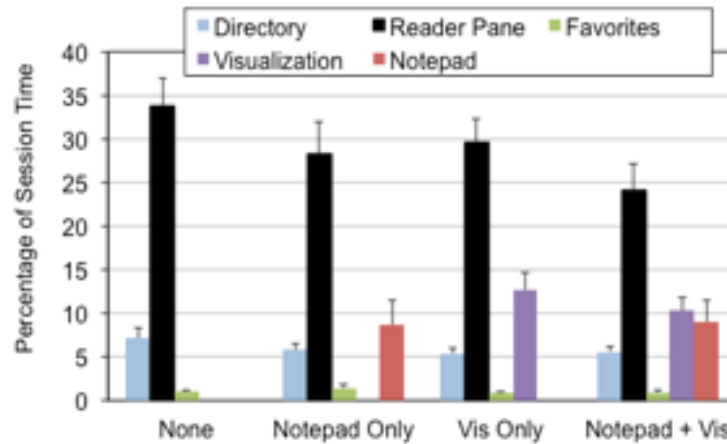


Figure 5. Percentage of time spent using the Directory, Reader pane, Favorites, Visualization (if available) and Notepad (if available). (Error bars represent standard errors of the mean.)

Having both the visualization and the notepad did not influence the amount of time people spent using either the visualization or the notepad. Two one-way ANOVAs comparing time using the visualization tool with and without the notepad and time using the notepad tool with and without the visualization showed no significant differences (both $F[1,19] < 1$, *n.s.*).

Although it seems from Figure 5 that the more tools available, the less time spent on the documents, this was only true for the notepad tool. I ran a 2 (visualization condition) by 2 (notepad condition) ANOVA using time on the documents as a dependent measure. There was no effect of visualization condition ($F[1,39] < 1$, *n.s.*) but a significant effect of notepad condition ($F[1,39] = 7.95$, $p < .01$). There was no interaction ($F[1,39] < 1$, *n.s.*).

The amount of time spent on the visualization had no impact on how many clues participants detected. I ran a one-way ANOVA using only participants who had the visualization tool. Notepad tool presence vs. absence was used as a between-subjects factor and percentage of time spent on the visualization was used as a covariate. Results showed that neither factor affected the number of clues detected ($F < 1$, n.s., for both main effects and the interaction). This suggests that some other factor of the visualization tool, rather than the time spent on it, helped participants in finding clues and better solving the serial killer task.

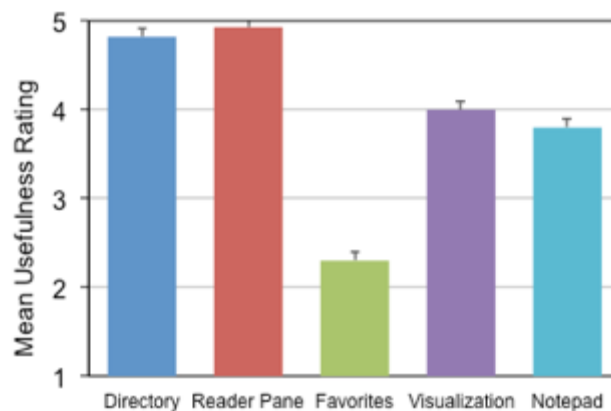


Figure 6. Mean ratings of tool usefulness. Higher values indicate greater usefulness.
(Error bars represent standard errors of the mean.)

Perceived tool usefulness

Finally, participants rated the usefulness of three tools common to all conditions (directory, reader pane, favorites) and the tools specific to their condition (notepad or visualization, if available) on 5-point scales. As shown in Figure 6, the directory and reader pane were rated as most useful, the notepad and visualization (when available) as somewhat less useful, and the favorites as least useful.

In addition to these findings, participants' open-ended responses in the survey shed light on the benefits and downsides of these tools for working on the analysis task and for the overall user experience. Consistent with the quantitative ratings, participants said they found value in the visualization for understanding how different pieces of information were related. For example, one participant said,

I was surprised, actually, how useful the visualization turned out to be. For just looking at one case, it isn't very useful, but it is useful for trying to find links between cases (P10, male, V)

At the same time, participants saw limitations in the visualization in its current form. For example, they pointed out that nodes could be linked based on randomly shared words between the documents, which could lead to depicting too many edges:

[In the visualization] there are misleading connections at times that are simply based on the words that are listed in groups of documents. (P39, male, VN)

The visualization was useful for some of the cases, but because it showed so many connections it seemed a little too imprecise to be extremely useful. (P20, female, V)

Similarly, participants saw value in the notepad tool. They liked the ability to collect, organize and revisit pieces of information in one place in a flexible, free-form structure:

I liked the notepad because it allowed me to organize with much more freedom in style I was not bounded by the confines of the program nearly as much. (P5, male, N)

I used the notepad a lot to gather my notes and see and reread information that I had highlighted. It was helpful to go back and read what I had written. (P11, female, VN)

At the same time, participants pointed to limitations in the notepad's editing functionality, lack of sketching capabilities, and annotations that were decontextualized from the documents from which they were taken:

Some sort of other features would have been nice (to help with organization).

Highlight, different colors, bolding, italicized, etc. (P24, female, N)

The notepad window was not useful because I found it more effective to view the important information in the context of the document where I found it, not by itself. (P23, male, N)

It would be nice to be able to scribble on or add comments to files (P10, male, V)

The design of the notepad as a simple text editor might have missed the richness of plain paper that can be freely scribbled upon. These comments point to ways in which a note-taking tool could be better designed and perhaps overcome the lack of support in performing the task.

In general, participants' ratings of the usefulness of the three system features common to all four conditions did not vary as a function of which other tools they had available to them. The

one exception was that participants without the notepad rated the favorites tool significantly more useful ($M=2.84$, $SD=1.38$) than participants with the notepad ($M=1.72$, $SD=1.13$); $F(1,39)=6.85$, $p=.01$. However, ratings of Favorites by participants in all conditions were quite low. In the open-ended survey responses, several participants mentioned that the data set was small enough and the directory was so well-organized that no additional organization of the documents in the Favorites was necessary.

Cognitive Load

Participants' ratings of their mental workload during the task indicated that they found it somewhat taxing, with means just over 4 on the seven-point scale (control: $M=4.08$, $SD=1.08$; V: $M=4.24$, $SD=.91$; N: $M=4.42$, $SD=.96$; VN: $M=4.56$, $SD=.80$). A 2 (visualization condition) by 2 (notepad condition) ANOVA on these scores revealed no significant effect of visualization tool ($F[1,39]<1$, $p=.62$) or notepad ($F[1,39]=1.22$, $p=.28$), and no interaction ($F[1,39]<1$, $p=.97$). However, greater percentage time spent using the notepad was significantly correlated with higher TLX scores ($r=.47$, $p<.05$). There was no relationship between time spent using the visualization tool and TLX scores ($r=-.001$, n.s.). These findings suggest that having both tools rather than a single tool did not create cognitive overload.

Discussion

In this chapter I first presented the design of the SAVANT system that takes a modular approach to supporting sensemaking in crime analysis. Individual features of SAVANT such as visualizations, notepads, and timelines, can be selectively turned on and off. Using this platform, I then conducted a controlled laboratory experiment looking at how two specific features of the

tool, a visualization tool and a notepad) facilitate or hinder the analysis task, how their users perceive them, and how I can learn from mistakes to provide analysts with better tools.

As H1 predicted, the visualization tool helped people identify and recall key clues that a serial killer was responsible for many of the homicides they were investigating and to identify the culprit. While it is unclear *how* the visualization tool led to improved performance, my observations of the experimental sessions suggest several ideas. Interacting with the visualization might have made it more obvious to participants that there were in fact connections among the documents from separate crime cases, and thus potentially a single underlying culprit that they should look for among the documents. In addition, the visualization may have given helpful starting points for the investigation, shaping the paths participants took through the documents instead of reading them in the order in which they were organized in the directory. Finally, the visualization might also have shaped the strategies participants used to investigate the homicides, for example skimming many documents linked in the visualization to identify underlying patterns, rather than reading each document in depth, a more cognitively taxing task.

Counter to H2, I found that the notepad tool provided no task performance benefits to my participants, alone or in combination with the visualization tool (in response to RQ), even though participants found it subjectively as useful as the visualization tool. This feature, like the visualization tool, underwent substantial preliminary testing and refinement prior to the study. Based on my findings, I have several thoughts about why the resulting feature provided less value than I had hoped. One possibility is that while people successfully used the notepad to organize their thoughts, as they noted in the open-ended responses, it slowed down their reading of the documents themselves, which led to no net benefit for analysis. Similar to other designs

(Piotech et al., 2006; Wright et al., 2006), I plan to explore ways in which a notetaking tool can be designed not only for collecting annotations, but also for flexibly organizing and linking them to each other and to other information objects.

Limitations and Further Directions

The initial SAVANT prototype focused on only some of the features that analysts would need for collecting data, identifying relationships among information entities, and making decisions. The notepad, for example, enabled capturing snippets of text from the documents and typing in free-form text. However, the current version of the tool lacks more extensive capabilities for diagramming relationships uncovered during the process of solving the crimes, which have been found to be beneficial for sensemaking (Piotech et al., 2006; Wright et al., 2006). Such a tool may be similar to scratch paper or a whiteboard, where the user can create diagrams by dragging information objects and creating new ones, drawing arrows and constructing connections, and highlighting objects they find important. Furthermore, unlike physical paper or whiteboard, this tool will have the power of connecting its content to the original dataset. In the next chapters, I will discuss ways in which an annotation and sketching tool can be designed to not only provide perceived value to users, but also improve their performance.

Another limitation of the current prototype is its focus on the individual analyst, with no means for sharing documents, notes, or states of the visualization with others. In Chapter 4, I will discuss how solo SAVANT was adapted to create a simpler collaborative version, offering analysts similar semblance to paper-based drawing by using shared free-form analytic space and digital Post-It Notes. The collaborative version of SAVANT, builds upon the same modular

nature of the solo version presented in this chapter, enabling customization of feature sets as identified to be necessary for the task at hand.

CHAPTER 4

COLLABORATIVE SAVANT: DESIGNING FOR IMPLICIT SHARING OF INFORMATION DURING CRIME ANALYSIS

As discussed in Chapter 1, the Boston Marathon Bombing suffered from lack of communication across team members, particularly between U.S. intelligence agencies and their Russian counterparts. Similarly, in 2007, Robert Pickton was convicted for six murders of women in British Columbia and connected to 24 others in the Vancouver region (Pickton Report, 2007). The Vancouver Police Department (VPD) came to suspect that the cases involved a serial killer, but they did not communicate this hypothesis to their cooperating partner, the Royal Canadian Mounted Police (RCMP). Furthermore, missing women reports filed with one agency were not shared with the other except when specific requests were made. These problems potentially delayed the investigation and led to more victims. In these and many similar cases a lack of communication led to a loss of important information that might even have prevented a crime.

In this chapter, I build on my work with the solo version of SAVANT (Chapter 3) by developing a collaborative version and using it to examine design solutions to the challenges of information sharing. Specifically, I examined the potential value of *implicit sharing of intermediate analytic insights*, which reduces the burden on the analyst to identify when and what information to share with partners. I focus on supporting the sharing of insights (e.g., hypotheses about who committed a crime, which analysts often find valuable for joint analysis (Convertino et al., 2011) rather than raw data and facts (e.g., clues from missing women reports)

because the latter is often subject to institutional policies (e.g., Cabrera, 2002), organizational policies and norms of sharing (Constant et al., 1994).

In the collaborative version of SAVANT, analysts make their own notes about insights from their own evidence. Depending on how the system is configured, those notes are either explicitly shared or automatically shared with their partners. When sharing is implicit (automatic), this removes the effort involved in assessing whether or not to share an insight and then explicitly placing it in a public workspace (Convertino et al., 2011; Hayne et al., 2011). My hypothesis is that because implicit sharing will reduce effort and increase sharing and awareness, analysts working collaboratively will perform better when implicit sharing is available than when it is not:

H1. Participants using implicit sharing of notes will perform better on a collaborative analysis task than participants without implicit sharing of notes.

Implicit sharing may also shift the value of certain elements of the analysis workspace. Individual features of analysis tools emphasize different aspects of sensemaking, and small changes to these features (such as making them shared vs. individual) can affect analysts' sensemaking strategies (Kang et al., 2011). A tool used to capture notes and insights may therefore become more valuable when it is shared with other analysts. Similarly, a workspace viewed only by a single analyst may be less valuable than one that multiple analysts can view. If analysts perceive that these tools are more valuable, they might use them more, interacting with their features and manipulating the data in them.

might reduce analysts' need to explicitly formulate messages and communicate information, thereby reducing their workload. On the other hand, shared workspaces might increase communication costs (Hayne et al., 2011). Seeing partners' activity might divert attention from one's own thoughts and increase the need for explicit discussion of process and data, especially when shared insights are connected to unshared data (Convertino et al., 2011). Since the direction of impact is unclear, I pose two research questions:

RQ1. How will implicit sharing of notes affect participants' cognitive workload?

RQ2: How will the availability of implicit sharing affect the amount of information exchanged via explicit channels?

Collaborative SAVANT

I adapted SAVANT system (Goyal et al., 2013) to make it suitable for collaborative analysis. SAVANT has two main components, the Document Space (Figure 7) and the Analysis Space (Figure 8). The Document Space has a number of features for data exploration and discovery. The document library and reader pane are for viewing and reading crime case reports, witness reports, testimonials, and other documents. A network diagram visualizes connections between documents based on commonly identified entities like persons, locations, and weapon types. The Document Space also provides a map of the area where crimes and events were reported and a timeline to assist in tracking events over time. Users can highlight and create annotations in the text of documents, locations on the map, and events in the timeline.

Such annotations automatically appear in the Analysis Space, an area for analysts to iteratively make and reorganize their notes until they see emerging patterns that lead to

hypotheses (Kang et al., 2011). These annotations are represented as digital *Stickies* (shaped as a Post-It note, a familiar metaphor preferred by analysts (Chin Jr. et al., 2009), which as in other analysis tools (Hayne et al., 2011; Kang et al., 2011; Pioch et al., 2006; Wright et al., 2006) are linked to the original document, map location, or timeline point where they were created. Stickies can also be created directly in the Analysis Space, unconnected to specific documents. Analysts can move Stickies around, connect Stickies together, or stack them in piles.

The Analysis Space supports collaboration through both explicit and implicit information sharing. For explicit sharing, a chat box at the bottom-left corner allows analysts to discuss their cases, data, and insights and to ask and answer questions. For implicit sharing, the Analysis Space shows other analysts' Stickies in real time as they are created and organized in the space. Stickies are color coded by the analyst who created them, but anyone can move, connect, or pile anyone's Stickies. Mouse cursors are independent of each other, while dependencies between Stickies are handled by the server on a first-come-first-serve basis. The server updates the interface every second.

Method

I designed a laboratory experiment in which two-person teams attempted to solve a set of crimes in a simulated geographically distributed environment. The task and materials were nearly identical to those used in Chapter 3. The crime cases were distributed between the partners, with a hidden serial killer that had to be identified. Half of the pairs worked on the task using an interface that provided implicit sharing of notes. The other half worked on the task using an interface without implicit sharing. I collected data via post-task surveys, participant reports, and computer logs.

I created two versions of SAVANT for this study. In the *implicit sharing* condition, Stickies in the Analysis Space were automatically shared as described above: there was no private workspace for analysis, only a public one. In the *no implicit sharing* condition, partners only see their own Stickies in the Analysis Space: there is no public workspace, only private ones for each analyst. The chat box is available in both conditions to support explicit sharing.

Participants

Participants consisted of 68 students at Cornell, a large northeastern university (22 female, 46 male; 85% U.S. born). Participants were assigned randomly into pairs, and each pair was randomly assigned to either the implicit sharing or the no implicit sharing condition.

Materials

The experimental materials were adapted from Balakrishnan et al. (Balakrishnan et al., 2008) and consisted of a set of practice materials and a primary task, similar to those used in Chapter 3. A practice booklet with a set of practice crime case documents introduced participants to the process of crime analysis and highlighted the importance of looking for motive, opportunity, and the lack of alibi. The primary task was created to be reasonable, but difficult, for novice analysts to complete in a limited time. The main task materials were a set of fictional homicide cases. There were six cold (unresolved) cases, and one current (active) case. Each of the cold cases included a single document with a summary of the crime: victim, time, method, and witness interviews. Four of these six cold cases were “serial killer” cases. These four had a similar crime pattern (e.g., killed by a blunt instrument). The active case consisted of nine documents: a cover sheet, coroner’s report, and witness and suspect interviews. Additional documents included three bus route timetables and a police department organization chart.

The documents were available through the SAVANT document library and were split between the two participants such that each had access to 3 cold cases (2 serial killer cases, 1 non-serial killer case) and 5 documents from the active case (both participants had access to the cover sheet). The additional documents were available to both participants. Overall, each participant had access to 13 documents, of which 6 were shared with the other participant and 7 were unique.

Twelve clues for detecting the serial killer were dispersed across the 20 documents with 40 suspects/witnesses, equally distributed between the two participants with four in common and four unique to each partner, following a hidden profile task paradigm (Stasser et al., 1985). The key clue to naming the killer was included in one of the witness reports of the active case, although the active case was not one of the serial killer cases. The task for this study was carefully designed to include data and aspects that are similar to real-world crime cases that remain unsolved, at a scale that could be analyzed in a one-hour session, and at a level of difficulty where many people are unable to solve the crime (Balakrishnan et al., 2008; Goyal et al., 2013).

Equipment

Two workstations (Intel Core i7 processor, 16 GB RAM) were connected to the Internet and ran SAVANT. Each was connected to two 25” monitors, the left showing the Document Space, and the right showing the Analysis Space. SAVANT logged keyboard and mouse activity as locally stored time-stamped CSV files. To simulate remote collaboration, the workstations were in separate cubicles to prevent eye contact and participants wore noise-cancelling headphones to prevent noises (e.g., keyboard and mouse clicks) from affecting each other.

Procedure

After being seated in separate cubicles, participants signed a written consent form, and read the training materials and performed the practice task individually for about 10 minutes. Participants then received a 10-minute tutorial on the SAVANT interface. The experimenter explained the different parts of SAVANT using example tasks that participants would perform.

Then, using SAVANT, participants worked as a team on the primary task to identify cases associated with a serial killer, name the serial killer, and find as many clues as possible in 60 minutes. At the end of the task, each participant received a paper report form at their workstation to fill out with name of the serial killer, associated cases, and the clues they could recall that would incriminate the killer. They then completed an online survey with questions about clue recognition, the utility of the interface, the collaboration experience, cognitive load, analytic ability (for control) and demographic information.

Measures

The following measures, similar to Study 1, were taken from data collected from system logs of interface use, post-task surveys and written reports.

Task performance

To address H1, which predicts that implicit sharing would improve task performance, I used three measures. The first two are based on participants' ability to remember clues pertaining to identifying the serial killer, and the third is whether they were able to name the killer.

Clue recall: At the end of the session, each participant wrote down as many clues as they could recall supporting their hypothesis about the serial killer. A participant's clue recall score

was the number of correct clues written down, similar to the measure used by Convertino et al. (2008, 2011).

Clue recognition: The post-task survey included multiple-choice questions, each related to one of the 10 clues hidden in the dataset. For example, “On the day of his wife’s murder, Ron Raffield claimed that A. He’d run into an old acquaintance on the bus. B. He’d been out of town on a business trip. C. He’d been tied up in a meeting all afternoon. D. He’d tried to call Darlene, but she never answered her cell. E. I do not know.” A participant’s score was the number of correct answers to these 12 questions.

Solving the case: At the end of the session, each participant wrote a report in which they were asked to name the serial killer. I counted this as binary variable: either the serial killer was identified (1) or not (0).

Perceived usefulness of SAVANT features

In order to answer H2a, I asked several questions probing participants’ evaluations of features of the SAVANT system in the post-task survey. This is similar to other studies that examined the usefulness of system features (Convertino et al., 2008; Wright et al., 2006).

Stickies: Four 5-point questions asked participants about the degree to which the Stickies promoted discussion, helped achieve understanding, and communicate ideas. For example, “The Stickies in Analysis Space helped me understand what my partner was thinking.” These four questions formed a reliable scale (Cronbach’s $\alpha=0.77$) and were averaged to create a measure of Stickies’ usefulness.

Analysis Space. Five 5-point questions asked about the degree to which the Analysis Space helped participants feel physically, cognitively, and emotionally closer to their partner,

helped them work with their partner, and helped them understand their partner's activities. These five questions formed a reliable scale (Cronbach's $\alpha=0.85$) and were averaged to measure Analysis Space usefulness.

Use of SAVANT features

In order to answer H2b, I used system logs to derive In order to answer H2b, I used system logs to derive measures of participants' actual use of features in the Analysis Space, including the number of *connections* they made between Stickies, the number of *piles* they created, and the overall number of *movements* (editing, adding, deleting, connecting, or piling) of Stickies.

In the implicit sharing condition participants could manipulate both their and their partner's Stickies, whereas in the non-implicit sharing condition each participant could only manipulate their own Stickies. Therefore, these three measures are at the pair level, aggregating both participants' actions in a session.

Team experience

The post-task survey contained ten survey questions about the quality of the collaboration (e.g., "It was easy to discuss the cases with my partner," "My partner and I agreed about how to solve the case"). These ten questions formed a reliable scale (Cronbach's $\alpha=0.84$) and were averaged to create a team experience score, to answer H3. This measure is similar to (Bier et al., 2010; Convertino et al., 2008) who used a post-task questionnaire to assess quality of communication within the group.

Cognitive load

In order to answer RQ1, the post-task survey contained five questions based on the NASA TLX (Hart et al., 1988) that asked participants to rate how mentally demanding, temporally demanding, effortful, and frustrating the task was, as well as their subjective performance. After inverting the performance question, these five responses formed a reliable scale (Cronbach's $\alpha=0.76$). Participants' responses were averaged to create one measure of cognitive load.

Explicit sharing

SAVANT logged the chat transcripts for each session, which were then cleaned to remove extraneous information like participant identification and timestamps. To answer RQ2, explicit sharing was measured at the pair level as the number of words exchanged in the chat box during a session. This is similar to (Hayne et al., 2011), who assessed the number of chat lines exchanged during the experimental session.

Results

I present the findings in six sections. First, I discuss effects of implicit sharing on my three task performance measures. I then consider how it affected subjective ratings of SAVANT features and their use, perceptions of team experience, cognitive load, and explicit information sharing

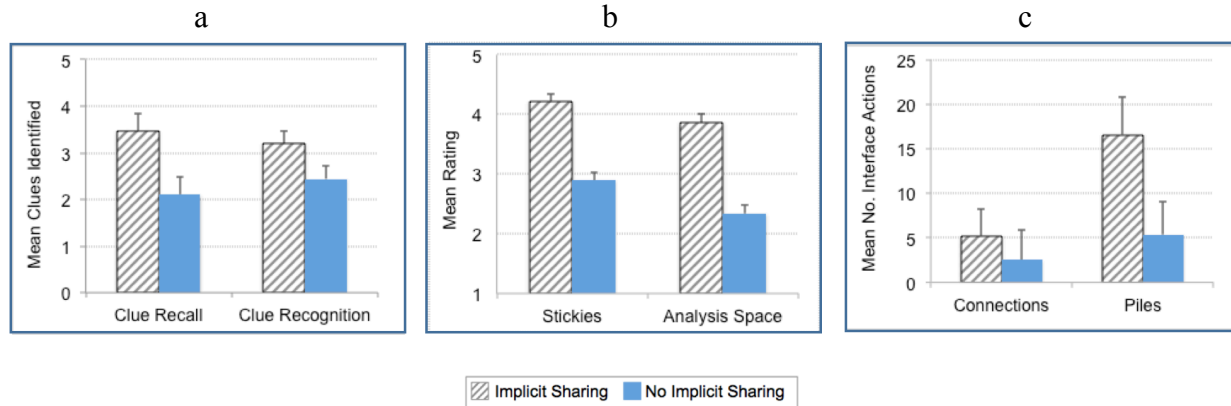


Figure 9. Results a) Task performance, (b) Perceived usefulness of Stickies and Analysis Space, and (c) Number of connections and piles made in a session, each by interface condition. Error bars represent standard errors of the mean.

Task performance

H1 proposed that pairs would perform better when implicit sharing was available than when it was not available. To test this hypothesis, I conducted mixed model ANOVAs, using clue recall and clue recognition as my dependent measures. In these models, participant nested within pair was a random factor and condition (implicit vs. no implicit sharing) was a fixed factor.

Clue recall. There was a significant effect of implicit vs. no implicit sharing on the number of clues participants recalled in the written report ($F[1, 66]=6.54, p=0.01$). As shown on the left side of Figure 9a, participants in the implicit sharing condition recalled more clues ($M=3.47, SE=0.37$) than those without implicit sharing ($M=2.11, SE=0.37$). Clue recall difference was also significant at the team level ($t[32]=2.03, p=0.05$), with teams with implicit sharing recalling more clues ($M=4.71, SE=0.58$) than teams without implicit sharing ($M=3.11, SE=0.52$). Given the large Cohen's d (3.68 for individuals, 2.90 for teams) and the fact that these

clues were buried in 20 documents with many information pieces, I regard this as a meaningful increase in clue recall, paralleling other work that has found increases in task-relevant recall in shared workspaces (Convertino et al., 2011 , McCarthy et al., 1991).

Clue recognition. The right-hand side of Figure 9a shows participants' performance on the multiple-choice clue recognition questions in the post-task survey. There were no statistically significant differences in clue recognition ($F[1, 66]=3.52, p=0.06$) between individuals in the implicit sharing ($M=3.20, SE=0.28$) and no implicit sharing conditions ($M=2.44, SE=0.28$). This was also consistent at the team level ($t[32]=0.80, p=0.42$) between teams with implicit sharing ($M=5.35, SE=0.41$) and no implicit sharing ($M=4.82, SE=0.51$).

Solving the case. I also examined whether interface condition affected the likelihood that participants could solve the crime. Since solving the case was a binary dependent variable, I ran a binomial logistic regression with condition as the independent variable and pair as the random effect variable. There was no significant difference between the implicit sharing condition ($M=0.62, SE=0.11$) and the no implicit sharing condition ($M=0.74, SE=0.01$; Wald Chi Square $[1, 68]=0.57, p=0.45$). Sharing knowledge manually did not improve answer accuracy in (Convertino et al., 2011) but sharing knowledge implicitly in a small experiment did increase answer accuracy in (Hayne et al., 2011).

Perceived usefulness of SAVANT features

H2a stated that participants would perceive SAVANT features as more valuable when the interface supported implicit sharing than when it did not. I analyzed participants' ratings of the usefulness of Stickies and of the Analysis Space using mixed model ANOVAs with participants

nested within pair as a random factor and interface condition (implicit sharing vs. no implicit sharing) as a fixed factor.

As shown in Figure 9b, H2a was supported. Participants in the implicit sharing condition viewed Stickies as more useful ($M=4.22$, $SE=0.12$) than those in the no implicit sharing condition ($M=2.90$, $SE=0.12$; $F[1, 66]=53.1$, $p<0.001$). Participants in the implicit sharing condition also rated the Analysis Space as more useful ($M=3.86$, $SE=0.14$) than those in the no implicit sharing condition ($M=2.34$, $SE=0.14$; $F[1, 32]=55.39$, $p<0.001$).

Use of SAVANT features

H2b predicted that the availability of implicit sharing would lead participants to interact more with Stickies in the Analysis Space than without implicit sharing. Using system logs, I counted the number of connections between Stickies, piles of Stickies, and overall Analysis Space manipulations that pairs made over the course of a session. Overall use of connections and piles was quite low and not normally distributed, so I did not perform ANOVAs on this data. Descriptive statistics (Figure 9c) suggest that in the implicit sharing condition, pairs created more connections ($M=5.24$, $SE=3.01$) and more piles ($M=16.59$, $SE=4.16$) than in the no implicit sharing condition (connections: $M=2.51$, $SE=3.38$; piles: $M=5.37$, $SE=3.63$). Participants in the implicit sharing condition also performed more total manipulations of Stickies ($M=1361.85$, $SE=149.02$) than in the no implicit sharing condition ($M=624.79$, $SE=81.95$).

Team experience

H3 predicted that participants would rate the quality of their collaborations with their partners higher when they could implicitly share information compared to when they could not. To test this hypothesis, participants' team experience scores were analyzed in a mixed model

ANOVA in which participant nested within pair was a random factor and condition (implicit sharing vs. no implicit sharing) was a fixed factor. H3 was not supported ($F[1, 36.30]=0.62$, $p=0.44$).

Cognitive workload

RQ1 asked whether cognitive workload would vary as a function of the presence or absence of implicit sharing. A mixed model ANOVA showed no significant difference between interface conditions ($F[1, 36.48]=2.49$, $p=0.12$); participants in the implicit sharing condition rated workload slightly but not significantly lower ($M=4.43$, $SE=0.18$) than in the no implicit sharing condition ($M=4.78$, $SE=0.18$).

Explicit sharing

RQ2 asked whether the availability of implicit sharing might change the amount of *explicit* sharing via the chat box. A one-way ANOVA was used to compare word counts at the pair level, using condition as the fixed factor. In contrast with (Doroudi et al., 2016 [54]), who found a significant increase in the amount of explicit chat communication in a shared versus a non-shared condition, I found no significant differences in word count ($F[1, 32]=2.11$, $p=0.16$) between the implicit ($M=782.59$, $SE=77.15$) and no implicit sharing conditions ($M=948.59$, $SE=84.40$).

Roles of implicit and explicit sharing

Participants' open-ended responses on the post-survey shed some light on just how implicit sharing was valuable and how it interacted with explicit sharing features. Several participants mentioned that implicitly shared Stickies helped them “*make connections*” and also

added value “by comparing information” or “cross-referencing information” visually between each other to promote awareness:

“The Stickies enabled a connection between my partner and I, I could see each other’s train of thoughts and methods of organization. I used the connecting lines for the Stickies to show myself and my partner the connections that I was seeing.”
(P27, Female).

Much of the value came from the combination of implicit and explicit sharing. For example, implicit sharing could reduce the need for explicit communication:

“The chat was easily the most helpful because it allowed me to communicate and tell each other specifics about the case. The Stickies were very useful also because they allowed me to make connections between the information I both had independent of talking with each other. [Stickies] allowed me to work more efficiently than wasting both of my time.” (P8, Male).

On the other hand, implicit sharing could also prompt explicit chat and sharing, when it revealed needs and gaps:

“I used the Stickies as jumping off points for conversations with my partner I would see her Sticky and then ask her to fill in some details that she may have skipped over since she had access to certain documents that I did not.” (P15, Female).

Finally, Stickies were intentionally designed to be free-form and open-ended, and participants did use them to separate aspects of the problem:

“I simply piled them together and placed them in strategic positions. I used two stickies sometimes for the same case. Each sticky would have another side of the case like emotional and the other would be factual.” (P61, Female)

“I took notes from the documents and highlighting the important parts created the stickies in the analysis space. I added my notes and thought as stickies in the analysis panes when I made connections between cases.” (P19, Female)

In addition to the quantitative results, these comments show *what* value participants found in implicit sharing, *how* it was used independently and in tandem with explicit forms of communication, and how it can be further improved. They demonstrate the power of implicit sharing to improve collaborative analysis without requiring partners to explicitly push or pull information by triggering understanding and insights on both sides, improving efficiency of conversation, and initiating explicit discussions.

Discussion

In this chapter, I developed a collaborative version of SAVANT and used it to test the value of implicit sharing of information among analysts. The findings show mixed results about the value of implicit sharing in terms of task performance (H1). Participants were better able to identify relevant clues in the data when implicit sharing was available, but were not better able to name the killer. The hidden profile nature of this dataset has been shown to make this task quite difficult (Balakrishnan et al., 2008), however, so it is promising that implicit sharing aided with

intermediate sensemaking processes.

Both participants' perceptions (H2a) and actual usage logs (H2b) of the Stickies and Analysis Space features of SAVANT suggest that these features were more valuable when implicitly shared, but that they did not increase cognitive workload (RQ1) or change the amount of explicit conversation about the case (RQ2). However, participants' appreciation of implicitly shared features in SAVANT did not carry over to improved perceptions of the team experience (H3), perhaps because factors such as task difficulty and distributed interaction harmed team dynamics as strongly as the tool helped them.

Based on these findings, I suggest a number of design implications that can further improve collaborative analysis performance and process. One important observation is that not all Stickies were created equal. my intent was that they would represent “insights” (as opposed to “facts”), but people appropriated the Stickies for a variety of purposes: tracking “emotional” versus “factual” sides of the case (P61); “highlighting important parts” of a case versus “making connections” (P19); “trains of thought” versus “methods of organization” (P27). Providing ways to distinguish between different kinds of analyst note (e.g., through color, font, size, shape) that help analysts bend the notes to their ways of thinking might encourage the sharing of different kinds of information that helps establish common ground (McCarthy et al., 1991). NLP techniques that distinguish document-based facts vs. inferential comments (Stoyanov et al., 2005; Wiebe et al., 2005) could be used to suggest categorizations, both making this feature smoother to use and encouraging analysts to be more aware of when they are making inferences.

Limitations and future directions

Additional research is required to determine the best way to implement implicit sharing, and to assess how the value of implicit sharing may change over longer time periods, with larger datasets, varied data types, teams, and analysis tasks on factors like cognitive-load etc. For example, implicit sharing may not scale to larger tasks if the amount of information to process starts to outweigh the advantages of awareness. Convertino et al.'s explicit moving of information from private to public workspaces (2011), Hayne et al.'s idea of parallel personal-but-visible-to-all workspaces (2011), and my choice of implicit sharing of all analytical activity are three points in what I see as a large design space for knowledge sharing.

Summary

The contribution of this chapter is two-fold a) showing how intermediate analytic artifacts like insights on a Post-It note when shared may offer advantages for task performance and collaborative team experience b) implicit sharing offers a complementary channel to explicit information sharing. In Chapter 5 I discuss the evolution of collaborative SAVANT into REFLECTIVA. REFLECTIVA builds upon SAVANT's finding that implicit information sharing is important and offers implicit sharing as default. REFLECTIVA also continues to leverage intermediate data analytic artifacts (like Post It Notes in SAVANT) and expands upon this notion by driving real-time visualizations based on such artifacts.

CHAPTER 5

REFLECTIVA: DESIGNING FOR SENSEMAKING TRANSLUCENCE

In the case of Boston Marathon Bombing introduced in Chapter 1, redditors implicated the wrong person and many other redditors confirmed this suspicion without fully verifying it, an effect known as cognitive tunneling: once people start down a path, they continue to seek evidence confirming that they are on the right track and ignore evidence that might suggest otherwise. Similarly, in March 2008, Demetrius Smith was charged for the murder of Robert Long. Even though Long was working with police as an informant and potential witness against his boss, Morales, police ignored Morales as a potential suspect even though he had both motive and opportunity. After serving a five-year prison sentence, Smith was exonerated and released due to evidence that some of the original testimony was racially biased.

Cognitive tunneling hindered the process of sensemaking in the Long murder case in two ways: First, the investigators should not only have collected evidence that confirmed their (wrong) hypothesis that Smith committed the crime, but also collected evidence that disconfirmed their hypothesis. Second, self-awareness of personal biases is hard. It is even harder in the process of complex sensemaking like crime analysis. In retrospect, awareness of biases might have afforded investigators the cognizance that their attention was prematurely focused on a single suspect instead of appropriately distributed across other suspects, including Morales. Thus, an absence of due process and transparency into one's own mental process enabled biased sensemaking.

Police Chief periodical reports that on average, 16 murders occur every day that might never be solved and their perpetrators never arrested because of reasons like confirmation biases

and groupthink (Police Chief Magazine, 2009). These issues may be exacerbated in cases where crime investigators in multiple agencies need to work together due to reduced information sharing and awareness across geographically distributed teams and investigating partners (Egger, 20020; Mentis et al., 2009). While the timely exchange of information is essential to successfully solving crimes, at the same time, information received from one analyst can unduly influence another's reasoning, resulting in cognitive tunneling.

This chapter builds upon the previous chapter by holding implicit sharing as the default while expanding upon the notion of using intermediate data analytic artifacts. In the current chapter, I focus on using such artifacts to create the notion of sensemaking translucence, or the process of making analysts more aware of their sensemaking processes. Sensemaking involves foraging for information pieces that could connect with each other, resulting in multiple initial hypotheses. These hypotheses are then closely synthesized to find evidence that confirms or disconfirms them, until an ultimate hypothesis remains (Pirolli & Card, 2005). Successful crime investigators pursue multiple suspects until they have sufficient information to rule out all but the culprit (Heuer, 1999). While the sensemaking process frequently goes wrong when information is not shared in a timely fashion, it can also go wrong when an analyst prematurely decides on a suspect without ruling out the others as in the case of Demetrius Smith.

To balance the need for information exchange with the goal of reducing cognitive biases, I developed REFLECTIVA, a sensemaking translucence interface that consists of two integrated parts: a *hypothesis window* that is intended to motivate explicit interchange of ideas about suspects' means, motives and alibis and a *suspect visualization* that provides automatic feedback on which suspects have been discussed based on the hypothesis window, a group chat window,

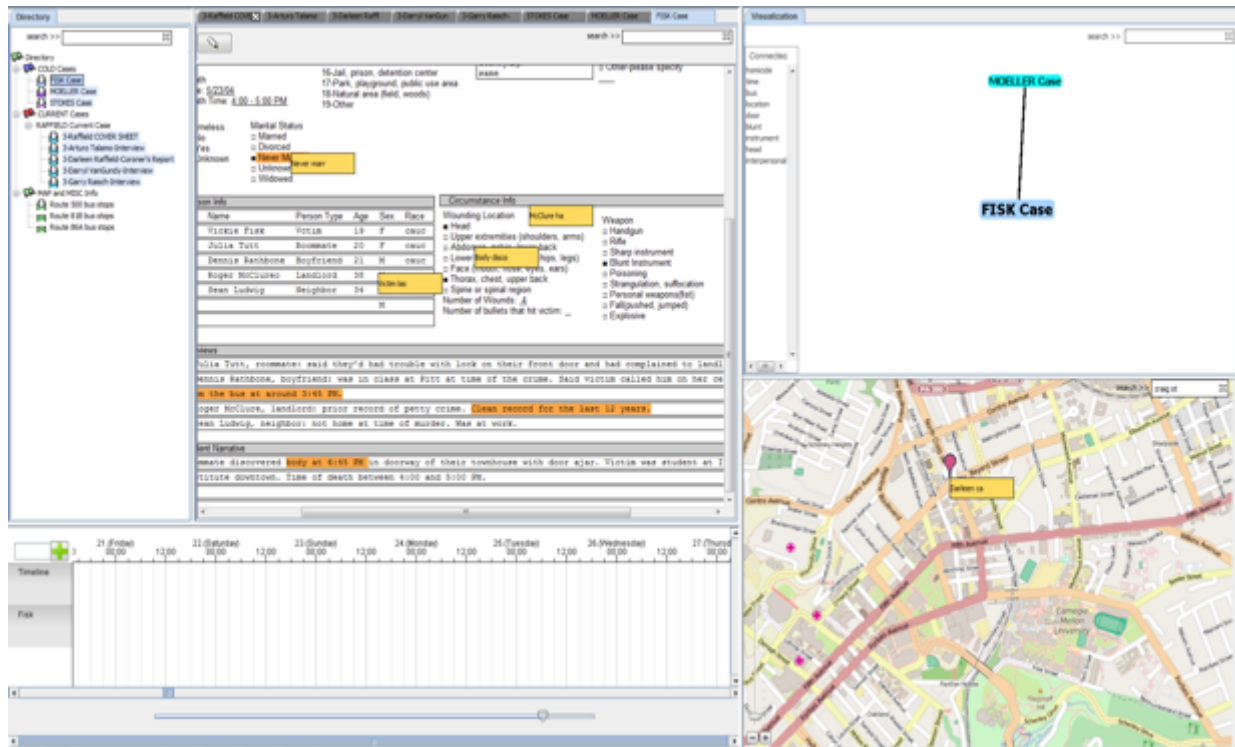


Figure 10. The REFLECTIVA Document Space showing (clockwise, from top-left) the directory of crime case documents, a tabbed reader pane for reading case documents, a visual graph of connections based on common entities in the dataset, a map to identify locations of crimes/events, and a timeline to track events

and a digital sticky note feature. The design of the suspect visualization is intended to provide awareness not only of those suspects that have been discussed but also of the idea that there might be other suspects out there that have yet to be discussed. I then examined the effects of the sensemaking translucence interface in a laboratory study in which pairs of remote participants role-played detectives collaborating to solve a serial killer crime (Goyal et al., 2016).

The *hypothesis window* is similar to Alternative Competing hypothesis (ACH) (Convertino et al., 2008), in which users explicitly share their hypotheses and evidence to maintain awareness of one another's insights and to develop a joint mental model of the case.

The hypothesis window is

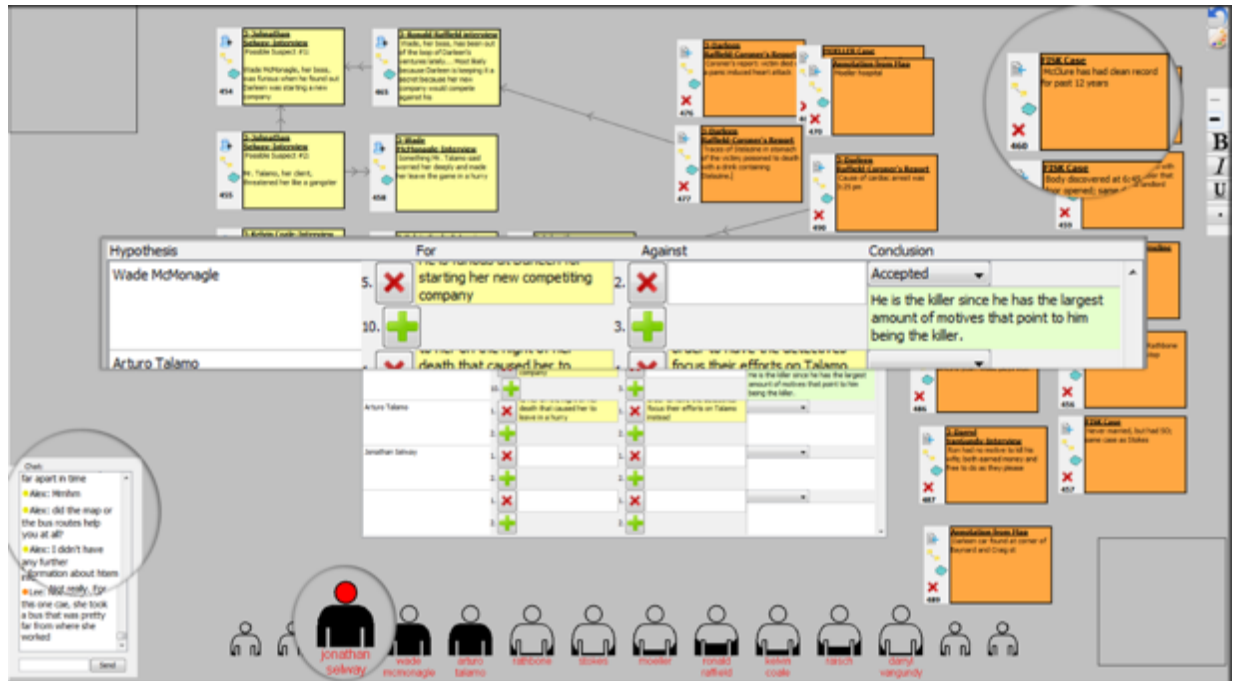


Figure 11. The REFLECTIVA Analysis Space showing (clockwise, from top-left) the chat for explicit sharing, connected stickies for Implicit Sharing, hypothesis window in the middle with columns to add new Hypotheses, confirming evidence, and disconfirming evidence for explicit Hypothesis Tracking, and Suspect Visualization at the bottom with 4 Avatars: Dennis Rathbone, Marilyn Stokes, Steve Gramming, and Lousie for Suspect Tracking. Note: Chat, Visualization, Sticky, and Hypothesis Window have been magnified to improve readability.

also designed to help reduce confirmation bias by including fields for reporting evidence that disconfirms each hypothesis (Goyal et al., 2016).

The *suspect visualization* depicts the joint attention paid to each suspect thus far in the analysis and encourages collaborators to distribute their attention across multiple suspects instead of focusing prematurely on a single suspect who might not be the actual culprit. The suspect visualization changes automatically as analysts mention suspects in their hypotheses, notes or chat conversations.

Study Hypotheses

The hypothesis window and suspect visualization are designed to be used in tandem, such that each new sharing of a hypothesis is associated with steps to assess the quality of that hypothesis (via fields in the hypothesis window) and steps to promote consideration of other possible hypotheses (via the suspect visualization). I thus tested a sensemaking translucence interface that contained these integrated features against an earlier version of the same tool that did not (Goyal et al., 2015). For the reasons outlined above, I predicted that the sensemaking translucence interface would improve pairs' crime-solving performance:

H1. Participants using a sensemaking translucence interface will perform better on a collaborative analysis task than participants using a standard interface.

I also reasoned that by enabling analysts with a better understanding of their partners' thoughts and activities, the sensemaking translucence interface would help analysts make appropriate decisions about their own activity (Froehlich et al., 2004) and that analysts would perceive the sensemaking translucence interface to be of more value for their work than the standard interface.

H2a. Participants using a sensemaking translucence interface will rate the usefulness of the tool higher than participants using a standard interface.

H2b. Participants using a sensemaking translucence interface will report higher level of activity than participants using a standard interface.

A sensemaking translucence interface also has the potential to improve the experience of working together. Awareness of other analysts' activities has been shown to help novice analysts get up to speed (Bier et al., 2010). In medical settings, implicitly shared awareness information can help establish common ground between clinical staff (Bardram et al., 2006; Paul et al., 2010) and lead to more positive perceptions of the process (Convertino et al., 2008). Since making sensemaking more transparent reduces uncertainty about the status of the task and reduces the need for verbal updates of status via the chat interface, I predicted:

H3. Participants using a sensemaking translucence interface will rate their collaborative experience higher than participants using a standard interface.

However, sensemaking translucence may also come with costs. Analysts may feel compelled to share preliminary thoughts, and read their partners' emergent hypotheses. This may increase the cognitive demand of the crimesolving task. On the other hand, by reducing the need for explicit verbal sharing of information, my interface may reduce the time and effort required for the task (Fisher et al., 2012; Weick et al., 1993). There is also a potential for the suspect visualization to be distracting. Since the direction of impact is unclear, I posed a research question:

RQ1. How will the sensemaking translucence interface affect participants' cognitive workload?

REFLECTIVA

REFLECTIVA, the tool used for this experiment is based on SAVANT from previous chapters. REFLECTIVA has two main components: a *Document Space* and an *Analysis Space*.

The *Document Space* (Figure 10) was identical for both the standard interface and the sensemaking translucence interface. Here, investigators could view their case documents, and highlight/annotate text in these documents. They could also view and manipulate a network diagram that showed connections between cases as calculated by TF/IDF on named entities, access and annotate Google Maps to mark crime-locations, and annotate a timeline to identify temporal patterns. The Document Space appeared on one of the analysts' two monitors.

A second monitor was used to present the *Analysis Space* (Figure 11). Two features of the Analysis Space were common to participants in both the Standard Interface and the sensemaking translucence interface: digital stickies and a chat box. Annotations created in the Document Space appeared automatically as digital stickies in the Analysis Space, where they could be moved, edited, connected using arrows, or piled atop one another to show relevance. This iterative reorganization of stickies supports analysts' processes of foraging and sensemaking (Chin Jr. et al., 2009; Convertino et al., 2009; Hayne et al., 2011; Pioch et al., 2006). The Analysis Space also included a standard chat box (Figure 11b, lower left).

The Analysis Space for participants in the sensemaking translucence interface included two additional features: a *Hypothesis Window*, and a *Suspect Visualization*. These two features are connected to each other, and enable sensemaking translucence in two different ways.

The *Hypothesis Window* (Figure 11, center) allows users to enter their emergent hypotheses manually, reflecting on their current cognitive state of sensemaking. This space also reminds users to add evidence that confirms and disconfirms these hypotheses, such that users can explicitly mark the status of each hypothesis as accepted, rejected, or needing more information. Entries (hypothesis, confirming/disconfirming evidence, status, and status related

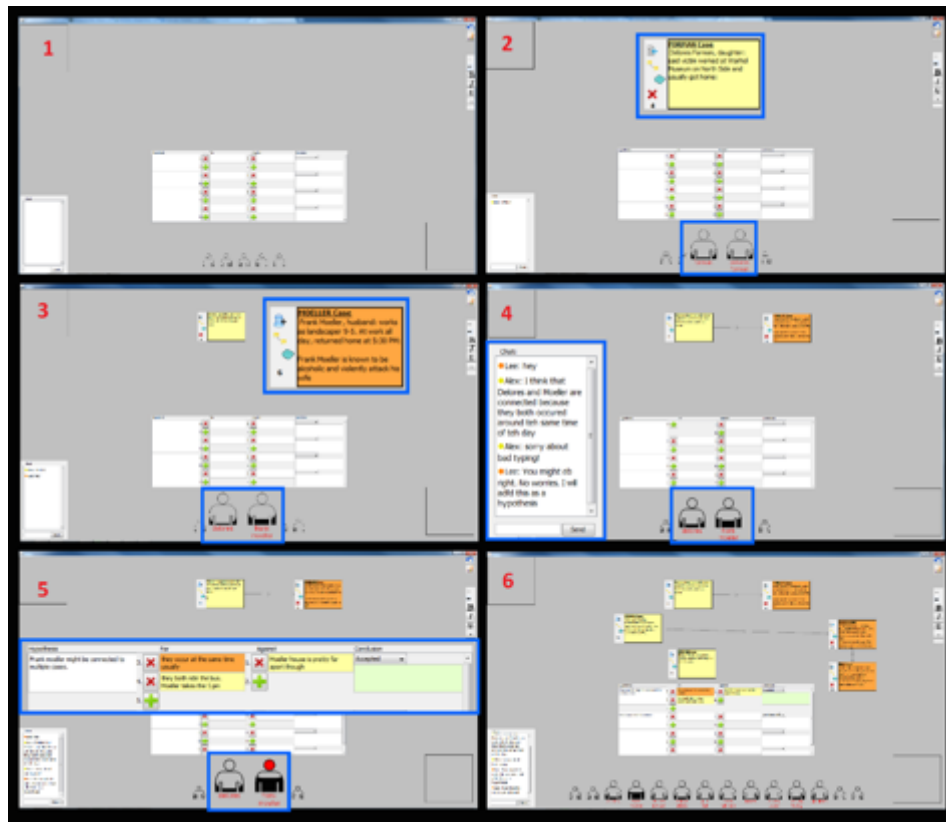


Figure 12. Sample sensemaking trajectory.

1. Sensemaking-Translucence reminds users to consider suspects by showing empty Avatars at the start
 2. Avatars are automatically populated by names detected from implicitly shared stickies.
 - 3, 4 & 5. Avatars show distribution of name-reference by getting darker for names mentioned in stickies, chat and hypothesis window. The last mentioned suspect in hypothesis window is marked red
 6. With use, visualization depicts distribution of attention at suspect level, based on explicit mentions.
- Note: Chat, Visualization, Sticky, and Hypothesis Window have been magnified to improve readability in 2, 3, 4 and 5. 1 and 6 represent non-magnified versions of the Sensemaking Translucence interface.

comments) were color coded to reflect each team member's contribution.

The *Suspect Visualization* was generated by the REFLECTIVA system in real time using Natural Language Processing of named entities. The system automatically identified named entities (names of persons only) in stickies, the chat conversation, and the hypothesis window. Each newly identified name was assigned an avatar. The visualization begins with four unnamed

avatars, suggesting that users should pursue names of potential suspects while the potential suspect-space is empty. As users share more suspect names in the Analysis Space, newly created named Avatars flanked by unnamed Avatars further remind users that there may be more suspects to discover. Furthermore, each time a name is mentioned in the Analysis Space, the associated Avatar darkens. This reflects the lack of non-equitable distribution of information sharing in the Analysis Space and supports suspect tracking.

Figure 12 shows one possible sensemaking trajectory where the two sensemaking translucence features facilitated the exchange of insights while simultaneously discouraging cognitive tunneling.

Method

Pairs collaborated to identify a pattern in a crime dataset similar to that used in the SAVANT studies reported in Chapters 3 and 4. They used a simulated geographically distributed environment. Pairs were randomly assigned to one of two interface conditions: standard interface or sensemaking translucence interface. The *Standard Interface* included a document space and an analysis space where users could share information using stickies and chat. The *Sensemaking Translucence Interface* allowed users to share information like the standard interface and further enabled partners to track the progress of their analysis by explicit hypothesis tracking and suspect tracking visualization. I measured task performance, perceptions of the interface, quality of the collaborative experience and cognitive load.

In summary, there were two different versions of the REFLECTIVA interface. In the standard interface condition, participants had the Document Space and an Analysis Space with stickies and chat box. In the sensemaking translucence interface condition, participants had the

Document Space and an Analysis Space that included the Hypothesis Window and Suspect Visualization in addition to stickies and a chat box.

Participants

Fifty participants participated in the experiment described as a “Solve Crimes Together Study” as 25 pairs. Of the 25 pairs, data for five pairs was discarded due to technical failures in Internet connectivity (4 pairs) and inconsistent instructions (1 pair). Finally, forty participants participated in the experiment (16 male, 24 female; 77.5% U.S. born; age range 18-28, median age approximately 21; 82.5% spoke English as first language). All students were undergraduate or graduate students at a large U.S. university. Participants were paid \$15 for their participation in the 1.5-hour experiment. Preliminary screening showed no significant demographic differences between participants in the two interface conditions.

Materials

Serial Killer Task. The task was based on the crime-solving paradigm presented in Chapter 3 and Appendix A. In this task, each participant is provided with a set of documents pertaining to 3 cold murder cases, half of the documents pertaining to a current murder case, bus route information, and maps of the areas of the crimes. In total, there were seven murders, with about 40 potential suspects, hidden in about 20 documents divided equally between the two participants.

The task required participants to share their information in order to connect 10 clues spread across the cold cases and two extra clues in the unrelated current case. This combination of clues indicated that a serial killer was responsible for four of the cold cases and revealed the identity of that serial killer. In previous studies this task has proven to be quite difficult for

participants such that the majority fail to identify the Serial Killer (e.g., Balakrishnan et al., 2008; Balakrishnan et al., 2010; Goyal et al., 2015; Goyal et al., 2013).

Post-task report form. After completing the task, participants were given individual paper report forms to complete. They were asked to provide the name of the serial killer, associated victims, and all clues that could incriminate the serial killer.

Post-task survey. An online post-task survey asked participants about their user experience and interface utility, collaboration experience, cognitive load (TLX), analytic ability, and demographic information. As described in more detail in the Measures section below, most questions were answered on 5 point Likert Scales.

Equipment

Two workstations (Intel Core i7 processor, 16 GB RAM) were connected to the Internet and ran as a web application, deployed on the university server. Each was connected to two 25” monitors, the left showing the Document Space, and the right showing the Analysis Space. To simulate remote collaboration, the workstations were in separate cubicles to prevent eye contact and participants wore noise-cancelling headphones that prevented them from hearing their partner’s speech or typing.

Procedure

Participants were seated apart at workstations such that they could not see each other or their partner’s workstations. The experimenter explained that they would be role-playing detectives on a homicide team. After the participants signed the written consent form, they received training about the importance of *motive*, *opportunity*, and *lack of alibi* in solving crime cases. For experimenter’s internal record keeping, participants with sensemaking translucent

interface condition were assigned numbering in hundreds (100 onwards) and with control conditions were assigned numbering in tens (1 onwards). Next, they performed a 10-minute practice task in which they identified motive, opportunity and (lack of) alibi in a laptop theft crime case.

Next, participants received the instructions for collaborating on the crime task: to work together as a team, share information, and find the name of the serial killer. They were also given a demo of the interface for their condition. Pairs were given 50 minutes to read through their documents, identify and share clues, brainstorm hypotheses, and identify the name of the serial killer. Upon completion of the task, they individually filled out the post-task report form and then the post-task survey.

Measures

I had two main sources of data: participants' final reports, and post-task survey results.

Task performance. Two measures for task performance were used, both based on the post task report form. Serial killer identification was a binary variable: 1 when correctly identified and 0 otherwise. Since this binary measure does not tell me how much progress a team had made in solving the case when the serial killer was not identified, I also used a clue recall score measured by the number of correct clues listed on the report form.

Usefulness of Analysis Space. Participants responded to multiple questions in the post-task survey about the usefulness of the Analysis Space for spreading their attention across multiple cases, generating hypotheses, and collaborating on the task. These measures are based on those from other similar studies (Balakrishnan et al., 2008; Goyal et al., 2015; Wright et al., 2006).

Focused attention activity: Five 5-point questions asked participants about the degree to which they interacted with the Analysis Space to pay attention to potential suspects, consider other alternative suspects, rule out suspects, track progress of suspects, and notice persons they did not pay enough attention to. For example, “I paid attention to number of potential suspects I considered in the Analysis Space”. These five questions formed a reliable scale (Chronbach’s $\alpha=.84$) and were averaged to create a measure of Focused Attention.

Hypothesis activity: Three 5-point questions asked participants about the degree to which the participants interacted with the Analysis Space to create hypotheses, confirm hypotheses and disconfirm hypotheses. These three questions formed a reliable scale (Chronbach’s $\alpha=.71$) and were averaged to create a measure of hypothesis Activity.

Analysis Space utility: Five 5-point questions asked participants about the degree to which the Analysis Space helped them discuss cases with their partner, understand what their partner was thinking, track progress, and made them feel cognitively, and emotionally closer to their partner. These five questions formed a reliable scale (Chronbach’s $\alpha=.88$) and were averaged to create a measure of Analysis Space Utility.

Team experience. The post-task survey contained ten survey questions about the quality of the collaboration (e.g., “It was easy to discuss the cases with my partner,” “My partner and I agreed about how to solve the case”). These ten questions formed a reliable scale (Cronbach’s $\alpha=.84$) and were averaged to create a team experience score, to answer H3. This measure is similar to (Chuah et al., 2003; Goyal et al., 2015) who used a post-task questionnaire to assess quality of communication within the group.

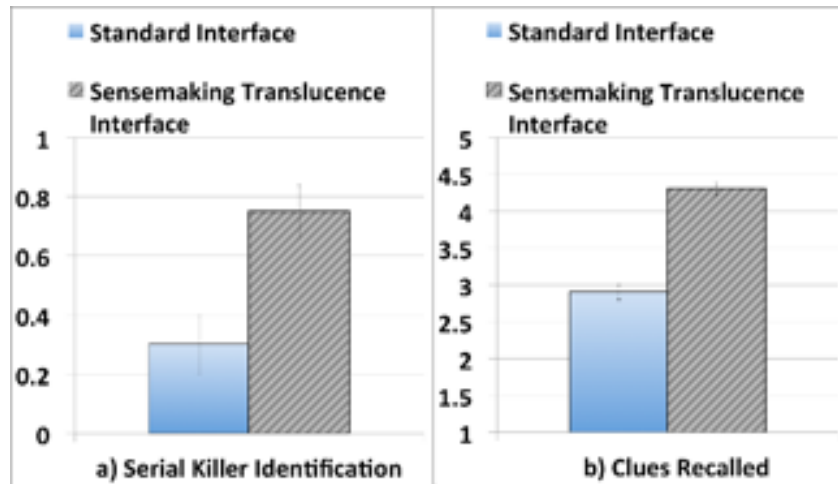


Figure 13. Serial killer identification and number of correct clues identified by interface condition.

Cognitive load. The post-task survey contained five questions based on the NASA TLX (Hart et al., 1988) that asked participants to rate how mentally demanding, temporally demanding, effortful, and frustrating the task was, as well as their subjective performance. After inverting the performance question, these five responses formed a reliable scale (Cronbach's $\alpha=.72$). Participants' responses were averaged to create one measure of cognitive load.

Results

I present the findings in four sections. First, I discuss the effects of sharing sensemaking translucence on my two task performance measures. I then consider how it affected subjective ratings of SAVANT features, subjective ratings of how participants interacted with SAVANT, perceptions of team experience, and cognitive load.

Task performance

H1 proposed that pairs would perform better when sensemaking translucence was available than when it was not available. To test this hypothesis, I conducted mixed model

ANOVAs, using clue recall and serial killer identification as my dependent measures. In these models, participant nested within pair was a random factor and interface condition (standard vs. sensemaking translucence) was a fixed factor.

Clue recall. There was a borderline significant effect of sensemaking translucence interface on the number of clues participants recalled in the written report ($F[1, 38]=3.80, p=.06$). As shown in a, participants using the sensemaking translucence interface recalled more clues ($M=4.3, SE=.47$) than those using the standard interface ($M=2.9, SE=.54$).

Serial Killer identification. b shows participants' performance at identifying the name of the serial killer. Participants were significantly more likely to identify the name of the serial killer correctly when using the sensemaking translucence interface ($M=.75, SE=.09$) than when using the standard interface ($M=.30, SE=.10; F[1, 38]=9.67, p=.004$).

Perception of usefulness of SAVANT features.

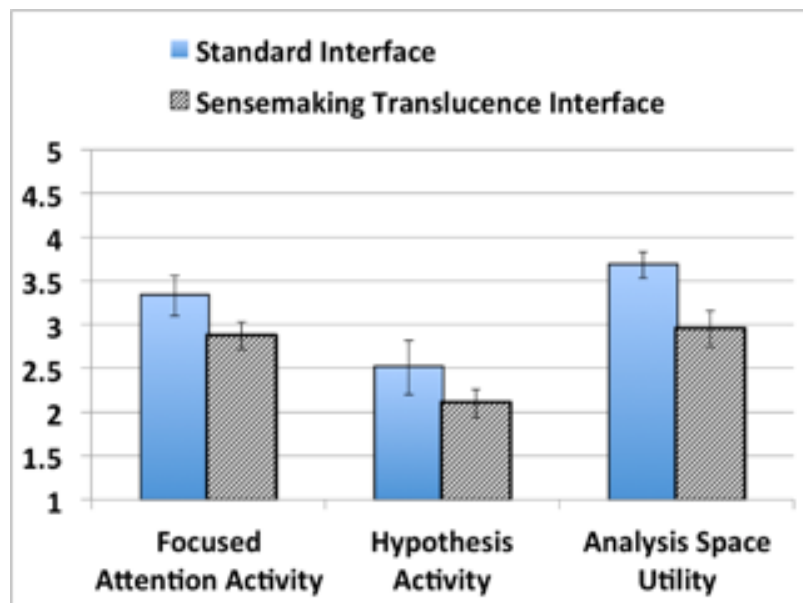


Figure 14. Perception of interface usefulness by interface condition

According to H2a, sensemaking translucence would be perceived as more valuable. I analyzed participants' self-reported ratings of the user activity with SAVANT's features using mixed model ANOVAs with participants nested within pair as a random factor and interface condition (sensemaking translucence vs. standard) as a fixed factor.

Both focused attention activity and hypothesis activity (left two graphs in Figure 14) show a negative trend and did not support H2b. Participants in the sensemaking translucence condition reported using the Analysis Space to pay attention to potential suspects less ($M=2.87$, $SE=.16$) than did those without sensemaking translucence ($M=3.33$, $SE=.23$; $F[1, 38]=2.56$, $p=.12$). Participants in the sensemaking translucence condition also reported creating, confirming, and disconfirming hypothesis lesser ($M=2.10$, $SE=.16$) than those with no task-monitoring ($M=2.51$, $SE=.31$; $F[1, 38]=1.39$, $p=.24$). These results are opposite to H2b.

Further, the participants rated Analysis Space to be of lower utility when sensemaking translucence was available and did not support H2a (right graph in). Participants in the non-sensemaking translucence condition reported Analysis Space to be significantly better at helping them discuss cases and feel closer to their partner ($M=3.68$, $SE=.15$) than when sensemaking translucence was available ($M=2.95$, $SE=.21$; $F[1, 38]=7.63$, $p=0.009$).

Team experience

H3 predicted that participants would rate the quality of their collaborations with their partners higher with sensemaking translucence compared to standard interface. To test this hypothesis, participants' team experience scores were analyzed in a mixed model ANOVA in which participant nested within pair was a random factor and condition (sensemaking

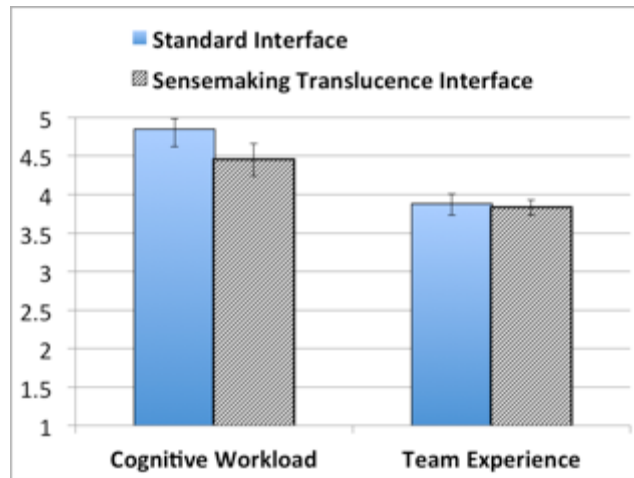


Figure 15. Self-reported workload and team experience by interface condition.

translucence vs. standard) was a fixed factor. H3 was not supported ($F[1, 38]=0.03$, $p=.84$, as shown in Figure 15.)

Cognitive workload

RQ1 asked whether cognitive workload would vary as a function of the presence or absence of sensemaking translucence. A mixed model ANOVA showed no significant difference between interface conditions ($F[1, 38]=1.55$, $p=.21$); participants with sensemaking translucence did not rate cognitive workload significantly lower ($M=4.45$, $SE=0.21$) than in the standard interface condition ($M=4.84$, $SE=0.22$). (See Figure 15.)

Roles of implicit and explicit sensemaking translucence

Participants' open-ended responses on the post-survey provided details about how the different features in Analysis Space were appropriated to collect clues, and solve the cases.

Hypothesis Window. First, several participants mentioned the interplay between the implicitly shared stickies and the Hypothesis Window. They referred to the bidirectional knowledge transfer between these two channels and showed how the two complemented each other:

“After deciding on an MO [Modus Operandus] for the serial killer in the hypothesis space, I moved that out of the window and onto a stickie.” – P111, female

“We used the analysis space to connect sticky notes and then form hypotheses based on the notes”...“We more sketched out ideas in the hypothesis space I think, after I had agreed on them” – P122, female

“The hypothesis window was very useful for synthesizing all the evidence I found in an easy-to-read window so that I could keep track of my findings. The stickies provided all the supporting evidence, but the hypothesis window summarized it for us.” – P106, male

Several participants also mentioned the partial use of the Hypothesis Window. Most participants created hypotheses towards the second-half of investigation, unless they were confident earlier on. Further, most participants also reported creating or supporting previously created hypothesis and shied away from adding evidences that disconfirmed hypothesis that had been decided upon previously with their partner:

“I created hypothesis for my cases with common MO. I confirmed the hypothesis made by me and the other person on my Current case. Didn't disconfirm the other person's hypothesis due to lack of detailed information about his/her cold cases.” – P103, female

“Used the hypothesis window by including supporting evidence only. I didn't write any of the hypotheses only my partner did. Didn't include any evidence to reject the hypothesis...” – P128, male

As is evident, much of the value of Hypothesis Window came from hypothesis reporting, evidence gathering, hypothesis confirmation, and combination with stickies in analysis space. Other participants shared their strategies of how they optimized use of the analysis space, and pointed to two distinct advantages of Hypothesis Window. First, it enables better organization of ideas than the free-form stickies. Second, it was used not to report the emergent hypothesis, but to report conclusions instead.

“I found that I was able to comprehend all the evidence best by looking at all of my stickies, setting up my hypotheses, and talking about what I thought. However, I think in order for stickies to be useful they have to be organized nicely. Sometimes when my partner's stickies were disorganized I had a hard time following her cases.” – P106, male

“Being able to chat was very important for talking about current ideas. Stickies helped to identify important things, but it was easy to overload on stickies, especially with 6 cases in one space. The hypothesis window was good for when I

already felt I had conclusions, but did not necessarily come in handy during the thought process. The visualization was good for seeing where I needed to do more work but not really good for focusing on people.” – P111, female

On the contrary, pairs without the Hypothesis Window also used both implicit and explicit channels, in agreement with previous research (Goyal et al., 2015):

“By using notes about each case and comparing them to one another.” – P14, female

“I wrote any details that could possibly mean any type of link (i.e.: someone worked with money, someone had a secret lover, someone worked in the same field as another victim).” P18, female

Suspect Visualization. Unlike the Hypothesis Window, participants reported using the Suspect Visualization for not just confirming, but also disconfirming their hypotheses. Several participants reported using the visualization to ensure that they did not focus attention on any particular case/suspect:

“Paid attention to the avatars mainly to make sure that I didn't concentrate on one case only” – P103, female

“I only paid attention to the avatars if I were talking too much about a certain person. I didn't think until the end to consider it for potential suspects.” – P115, male

Further, the visualization helped indicate lack of sufficient information about potential suspects, so that they could not be discounted. So, instead of removing potential suspects, visualization helped users to rule in potential suspects:

“I didn't rule out suspects due to the visualization but I did use the visualization to see if I needed more information on a suspect.” – P111, female

Avatars in the visualization also served to imply whether the pairs shared common ground about the potential correct answer by showing how much the collaborators were referring to any particular suspect together:

“I used the avatars to let me know that there was a certainty that my partner and I were on the same page about suspecting someone for the crime” – P109, male

For some others, the visualization either served as tool for confirming their “hunch” or had no effect on their sensemaking process:

“I hardly glanced at it. There were too many people for me to start accounting for them all. I only started looking into a suspect when I began noticing where they fit into the "story".” – P106, male

“Mentioning the name didn't really change my investigation of them or others didn't affect my work didn't rule anyone out” – P104, female

A few participants reported visualization blindness because visualizing suspects from previously reviewed cases could be irrelevant to newer threads of investigation.

“There were too many avatars for me to keep track of everyone, and I were on a time crunch. I don't think it is worth investigating someone more just because I haven't talked about them much. If they have a solid alibi, or are involved in a case that has nothing in common with my case, they are not as important as the person who is mentioned across several files.” – P106, male

“I didn't really use the Avatars and visualization at the bottom since I felt that it didn't really help and that it was there just as a distraction” – P110, female

On the other hand, other participants found value in the visualization by optimizing strategies rather than completely ignoring it. Some participants did not view the Avatars unless they visualized sensemaking translucence on the “local” current thread of investigation instead of “global” sensemaking translucence:

“I didn't use the avatars if they did not have to do with the case I was working on at the time. When I made new notes about small details, it added every single name I mentioned although it didn't have to do with another case I was working on.” – P109, male

Others found value in the visualization as sources for new investigation-threads and hypothesis based on the names, or how relevant these names seemed to be potential suspects:

“The avatars in the beginning were adding up very quickly. Not until two-thirds of the way into the time did certain avatars seem like true suspects to consider. And only then did my partner and I start developing hypotheses.” – P128, male

In addition to the quantitative results, these comments show how the Hypothesis Window was viewed as less useful than I expected. Even though participants liked its structured organization, they used it to primarily report and confirm conclusions instead of reporting emerging hypothesis and then working to confirm and disconfirm them. The Suspect Visualization was viewed as useful to identify potential suspects and validate common-ground. While both sensemaking translucence features connect with communication channels in the Analysis Space, the visualization improved collaborative analysis by creating common ground and initiating new threads of investigation.

Discussion

I developed REFLECTIVA to examine the value of sensemaking translucence in collaborative sensemaking and then evaluated it in a distributed synchronous collaborative crime-solving task. Consistent with H1, I found that pairs of analysts using an interface that provided sensemaking translucence identified significantly more clues relevant to solving the case and identified the culprit a significantly greater proportion of the time than pairs using the standard interface with no sensemaking translucence features. However, inconsistent with my other hypotheses, pairs using the sensemaking translucence interface rated it as less helpful than pairs using the standard interface in terms of providing support for analysis, aiding hypothesis generation, and helping them pay attention to multiple suspects.

There could be multiple reasons for this mismatch between the objective value of the tool for task performance and users' subjective perceptions of its value. First, participants may have been uncomfortable with the amount of explicit sharing of information and insights required in the sensemaking translucence interface even though this greater explicitness helped them solve

the case. The sensemaking translucence interface required explicit actions by the users in terms of when and what to write into the Hypothesis Window. This tension between explicit sharing for sensemaking translucence vs. ease of use could explain lower subjective ratings of the sensemaking translucence interface. Previous work shows that implicit sharing leads to significantly better user experience compared to when explicit actions are required by the collaborators (Goyal et al., 2015).

Second, participants had to spread their attention across far more features and information made available by these features in sensemaking translucence over the standard feature, even though this information helped them identify potential suspects. As P110 and P106 point out, participants using the sensemaking translucence interface had to distribute their attention between stickies, chat, an organized hypothesis window, and the suspect visualization whereas participants using the standard interface had only to attend to the stickies and chat. While the former did not lead to a significantly higher level of self-reported cognitive load, it may be that at some level participants reacted negatively to having to manage so many different features at once, resulting in lower reported perceived utility.

Third, despite the obvious gains in task performance, tools like my hypothesis window that enforce rigid structure are often perceived to be of low utility in collaborative sensemaking tasks. For example, Convertino et al. (2008) found that almost half of their participants wanted more control of how information was displayed in a matrix similar to SAVANT's Hypothesis Window. As P122 and P106 point out, even though the structured hypothesis window provided a strict organization to conclusions, the unstructured nature of the stickies allowed them to sketch out ideas prior to inserting their conclusions in the hypothesis window. However, previous

research indicates that despite the lack of *perceived* utility, rigid structure improves task performance in decision-making groups (Mentis et al., 2009).

In addition to the incongruence between my results for task performance and perceived usability, I also found that participants appropriated the sensemaking features differently because the two features were designed to support different parts of the foraging and the sensemaking process. While the Hypothesis Window was designed to encourage sensemaking through explicit confirmation/disconfirmation of partners' insights, the Suspect Visualization was designed to discourage focusing attention on any particular suspect early in the sensemaking process.

The Hypothesis Window was appropriated as a summarizing tool. One of the reasons could be that the participants may have to balance between the immediacy of hypothesis/evidence reporting with the perceived benefit of leaving their current sensemaking loop to do so. Some participants (P103 and P128) mentioned that they filled in the Hypothesis Window only towards the end of the process, rather than in an ongoing fashion as intended. This suggests a need to pay greater attention to the temporal nature of people's use of sensemaking translucence features. This connects well with Reddy's discussion of rhythms of work in information seeking (Reddy et al., 2002), which suggests that individuals' actions are dependent not just on immediacy but also on when it would be most beneficial to perform them in their work.

The Suspect Visualization was appropriated for foraging to rule in potential suspects. One of the reasons could be that the visualization affords an overall view of sensemaking across the 6 cases, instead of specifically supporting sensemaking of a single case. Users pointed to sensemaking as an act of pursuing multiple threads of investigation: global and local. As

suggested by some participants (P106 and P109), visualizing the “average” global attention-spread prevents depicting the attention-spread at a local case level, possibly leading them to view the Suspect Visualization as “distracting” when a pattern of a serial nature is not yet evident.

Limitations and Future Directions

This study also focused on two possible designs in a large design space for collaborative analysis tools, and it studied the impact of using the two features together instead of assessing the cost vs. benefit of each feature individually. More research will be needed to determine the best possible way to implement these features in collaborative sensemaking tools. Further, while this study focused on a study between a pair of collaborators, future research is needed to understand the impact of team size, and scaling across multiple collaborators (Goyal et al., 2013) on designing sensemaking translucence features.

One particular challenge might be the scalability of sensemaking translucence. With increasing time and complexity of datasets, or with increasing team size, explicitly sharing at a suspect and hypothesis level could lead to information overload. At the same time, however, it might provide data provenance that is often lost in sensemaking. Further, domains like crime-solving/medical-sensemaking etc. are constrained in the extent to which certain types of information can be shared due to organizational privacy laws. I see this as an open design space, where designers could consider single or multiple means of visualizing the sensemaking translucence.

While this work demonstrates the value of sensemaking translucence, much more work is needed to determine the best way(s) to implement it. For example, the tension between non-premature hypothesis sharing and timely information sharing should be reduced. I see this as a

spectrum where designers could design for fully automated hypothesis generation or require limited input from users to verify generated hypothesis, or any solution in between. Designers could push natural language processing further to identify not just the named-entities (suspects) but the explicitly shared hypothesis as well. Identified hypotheses may be implicitly shared using the hypotheses window, reducing the redundancy and improving sensemaking translucence. Improved sensemaking translucence without explicit sharing would reduce workload and will make user attention available for the process of sensemaking itself.

Providing interfaces that free up user attention could also help improve subjective user experience while maintaining high task performance. Connecting sensemaking translucence more closely to the artifacts of sensemaking itself could also reduce this tension between user-experience and task performance. The Hypothesis Window, and Suspect Visualization could be better integrated with the stickies to further accrue the advantages of implicit sharing. For example, evidence that disconfirms any hypothesis could be highlighted for closer inspection. Eventually, collaborators could leverage each other's insights as recommended by the system while they manage their limited attention under strict time pressure.

Future designs would also benefit from greater attention to how sensemaking features might support the different phases of sensemaking work itself. One design goal could be to remind users to interact with sensemaking translucence features more during the process of sensemaking. Future tools could match natural language processing on the chat transcripts with machine learning on user activity for behavior recognition. Recognizing activities correlated with task success vs. failure could help customize the tool usage, for instance by recommending to users that they need to pursue disconfirmation of existing hypothesis or that they should

distribute their attention across suspects. Detecting deviations from known successful behaviors and persuading investigators to reduce biases could potentially reduce the number of unsolved cases (Police Chief Magazine, 2009).

Finally, more closely integrated sensemaking translucence features could help reduce workload. In my design, integration was unidirectional such that the Hypothesis Window drove the visualization but not vice versa. It is also possible for the visualization to trigger new hypotheses about potential criminals. Implicit hypotheses (including location, timeline, alibi etc.) could be generated based on notes gathered for each suspect. This might aid sensemaking translucence by offering an alternative bottom-up view.

Summary

To summarize, the contributions of this chapter were a) introducing the notion of sensemaking translucence as a metaphor to overcome cognitive tunneling, a challenge I have seen surface multiple times b) leveraging intermedia data analytic artifacts to provide feedback to the teams about their progress and process. While REFLECTIVA affords the collaborators to collaborate and minimize cognitive tunneling, collaborators must still manage the collaboration process itself. For example, the collaborators must still manage incoming requests for information and information about new findings and insights, while managing their own iterative sensemaking process. Collaborative sensemaking process, hence, requires not just REFLECTIVA's anti-cognitive tunneling but also a way to manage and respond to such requests. In the next chapter, I will discuss the findings of the chapters 3-5.

CHAPTER 6

GENERAL DISCUSSION

The goal of this dissertation has been the creation and evaluation of novel designs for tools to support solo and collaborative sensemaking. The dissertation was organized into chapters that highlighted a novel tool design coupled with a study of its value for remotely located analysts. Each new tool aims to improve on the previous one by facilitating additional aspects of the analytic process.

One key observation in my work is that not all the design features are equally valuable. While Visualization and Note-taking features are individually important, the presence of both simultaneously has an adverse effect on task performance. Similarly, implicit and explicit sharing are both important. However, one cannot be discounted for the other. Each feature affords different kinds of information sharing.

A second key observation is that there is a tension between task performance and user-experience when users are exposed to reflexive visualizations. Tool designers may have to decide which of these they want to improve in critical situations like crime-solving.

A third key observation is that, despite the computational progress made by computer science, designing for sensemaking performed by analysts remains challenging because of the “human in the loop”. Sensemaking tools may be algorithmically advanced. However, unless our designs account for the usability, social and cognitive processes, no one may use them. In the rest of this chapter, I will highlight key results from the preceding chapters and then discuss how they contribute to the understanding of and resolution of this tension.

Summary of Results

In Chapter 3, I discuss how we should move away from one design fits all approach and focus on designing modular tools that have features which should be tested for their utility, experience, and interaction with other features before exposing a fully feature laden complex tool to analysts who might feel overwhelmed and not use it. I designed the SAVANT prototype, which allows researchers or analysts to selectively turn features on and off depending on their needs. In a laboratory study using this prototype, I discovered that different features impact analytic process differently, and their designs interact with each other – sometime even reducing the performance.

In Chapter 4, created a collaborative version of SAVANT and tested whether implicit sharing can enrich explicit sharing when analysts have to share information while working collaboratively across distance. In a lab experiment, I tested implicit sharing vs. no-implicit sharing and discovered that despite the danger of higher information overload, the automated sharing of insights implicitly did indeed improve both the task performance and team experience. However, not all implicit sharing is the same. For example, sharing of insights is different that sharing raw facts etc. I also discovered that automatically shared information has the power to confirm initial hunches held by the users and thus cognitively tunnel users' initial hypothesis.

In Chapter 5, I introduce REFLECTIVA, a tool iteratively designed based on SAVANT that includes additional features to create sensemaking translucence – automatically sharing team's focus of attention to highlight and expand the potential solution space and to disconfirm suggested hypotheses. Sensemaking translucence, I argued, could help reduce cognitive tunnelling by making analysts aware of their cognitive processes. In a lab experiment testing

REFLECTIVA against SAVANT, I found that while REFLECTIVA did improve task performance significantly over SAVANT, users still preferred SAVANT. It is possible REFLECTIVA pushed them to think at the team level and not at the individual level, or it reminded them to step out of their comfort zone and consider alternative solutions. This highlights a challenging decision for designers: design for performance or design for user experience.

In the next section, I will briefly reflect upon my findings presented in the previous chapters to identify opportunities for future designers designing for collaborative sensemaking

Design Implications

Pros and cons of different design decisions

Early on in my work, I discovered that different features afford differently abled sensemaking processes. For example, in an hour-long task, while notepad's presence decreased task-performance, longer sensemaking tasks might benefit from the notepad's provenance. Similarly, while a 2-member team might not overload each other with implicitly shared insights, a larger team or a longer task might generate significant intermediate analytic artefacts to overload the users, potentially reducing the advantages of implicit sharing. In short, while previous tools have been tested in their entirety, based on my findings I would advocate for modular tools that could be customized and personalized to the task at hand.

For example, when finding insights the tool could potentially minimize features not needed for the task. The document space could be expanded, revealing more features in the document space that might make the insight finding task easier. Such features may include, running NLP in real-time to find unique or common words in a particular document across the

dataset. Similarly, when analysing insights previously discovered, the analysis space could be expanded conversely revealing more interesting features. Such features may include offering new colors to attribute to different cases/stories etc.

Further, I discovered that evaluating these tools with human subjects reveals insights that would be missed, if human-centered design approach is ignored. For example, SAVANT's Visualization-only condition performed shockingly better than when both Visualization and Notepad were present. Similarly, Implicit sharing only helped enough to find more clues but led to cognitive tunnelling. REFLECTIVA's Hypothesis Window was used far more for record-keeping than for analysis suggesting that users preferred unstructured analysis. Only through the iterative human-centred design and evaluation could the benefits of every design decision be independently tested. Based on my findings, I advocate for future tools to evaluate every design decision independently and together prior to their adoption.

Role of technology in Sensemaking tool design

During the research presented in this dissertation, the technological landscapes have changed, and continue to change. We have moved from “Minority-Report”esque world to citizen scientists on Reddit Bureau of Investigation to predictive policing that helps allocate social services by leveraging AI to predict when/where future crisis is likely to occur. The role of technology and human is changing in critical sensemaking processes. Is technology aiding us in identifying criminals or are we cleaning up data to aid technology catch criminals? Is a self-driving car better at analysing who dies in a crash, or will it better for humans to decide which one of their own species dies? Which one is better, is a philosophical question beyond the scope of this dissertation.

However, I will focus on two technological advancements, intrinsically tied to this dissertation. Whether a human or a robot or a combination decides to solve crimes in the future, leveraging intermediate data analytic artefacts would be invaluable. Such artefacts would be necessary to expand each other's potential solution space, increasing the possibility of identifying globally optimal solution as opposed to a locally optimal solution. In SAVANT, such artefacts were the Post-It Notes in raw. In REFLECTIVA, these notes were processed into Visualizations. However, future technologies can leverage other intermediate analytic artefacts as well, including but not limited to identifying Hypothesis and processing evidences automatically to associate with confirming or disconfirming these Hypothesis. Once identified and tied to evidences, analysts could use these to rank how well a hypothesis is likely to hold.

Similarly, as we amass more data about judgements in crime-solving, we could potentially leverage technologies like deep learning. However, our crime datasets are not clean. Human behaviour induced biases is intertwined with these datasets. As we discover increasingly wrongful incriminations every passing year, it becomes even more imperative to remain vigilant about use of such datasets as predictors or learning sources for AI based technologies. Lack of vigilance, might lead to creating algorithms that learn to produce biased results based on past results. On the one hand, technology like REFLECTIVA can show us our biases. On the other hand, crowd-workers can clean up bad data or even perform analysis together, feeding into the machines. I am hopeful that through a combination of approaches, complex sensemaking tasks will not remain complex forever.

Automated documentation and credits

SAVANT and REFLECTIVA tools offer provenance of sensemaking process, the contributions of the collaborators, and enables story telling of how sensemaking transpired through usage log analysis. However, they also might seem constraining as they do not allow sensemaking efforts made beyond these tools to be reflected or captured within these systems. For example, analysts may use pen and paper or other tools like Excel etc. to perform analysis. While such tools offer simplicity and ease of use, SAVANT and REFLECTIVA offer the advantage of documented proof about who shared what insights and who was responsible for leveraging those insights.

Such documentation has been shown to be important when functioning in meritocratic organizations that credit analysts based on their performance. 9/11 Reports found that such systems prevent analysts from freely sharing information as the career progression is tied to them solving cases. Systems like SAVANT and REFLECTIVA when further developed may offer a documentation perspective to giving credit to whom it is due for sharing and leveraging insights together. Providing such data provenance might even encourage the analysts to use such tools more, and to share more freely.

Privacy and implicit sharing

While SAVANT and REFLECTIVA were designed to enable full sharing implicitly, the existing tools do not enable such level of sharing. The present-day tools enable analysts to create local notes and shoeboxes where important information may be saved for sensemaking task at hand or potential future sensemaking tasks. Analysts may then choose to share parts or whole of these artifacts with people they believe it would be advantageous to share with. Solo SAVANT

did indeed have a shoebox/bookmarking facility but it was removed for the collaborative version because retired analysts continue to argue that creating such local silos prevent information dissemination and knowledge sharing, leading to crime failures (Heuer, 1999).

Alternatively, crime investigation agencies also function in a hierarchical manner where access to information is determined by the role and relevance. Such organizational structures impose further limitations to information sharing. While privacy and security of data is necessary, future designers of sensemaking tools will have to find the balance between privacy and implicitly sharing all the notes. One may also argue that due to automated implicit sharing, analysts may choose to create lesser notes in danger of over sharing. In SAVANT, I found that to be quite opposite. Far more notes, connections, and piles were created during implicit sharing.

Finally, as a designer, I believe in transparency and such value system is represented in my design of SAVANT and REFLECTIVA. I believe that information, when shared freely, across the agencies and members can help solve crimes that have suffered from knowledge that was confined to silos. However, future research is needed to further understand the balance between implicit sharing and privacy by utilizing other potential value systems that might value privacy over transparency, for example.

General Limitations

There are a number of general limitations of this dissertation research, that highlight where future research may be pursued. Methodologically, Chapters 3-5 have focussed on using lab-experiments as a controlled setting to identify the impact of each design decision on task-performance and user experience. My findings agree with findings from previous literature about sensemaking and remote team-work. However, owing to the nature of the method, in-the-wild

studies might illustrate even more complex processes at work. In future, deploying secure systems similar to presented in this dissertation to understand how collaborative sensemaking may be improved with participants who are significantly impacted by the performance should be pursued.

Secondly, the experimental design itself focuses on 1-2 team members who are non-experts working together on an hour-long task. The results of this dissertation would help design tools for collaborating pairs performing a focused short-lived task, like solving a puzzle etc. Future work is needed to generalize the findings to experts working on longer tasks together in bigger teams. For example, the nature and volume of the dataset itself will change, affecting the use of different tools and design of such tools. While analysts would still potentially chunk to shoebox and iteratively access the smaller shoe-boxed information, SAVANT/REFLECTIVA might be unable to afford sufficient space for analysis. Large wall displays and VR might instead afford larger spaces to overcome this limitation.

Third, the presented setup for the lab experiments included no access to pen for annotations. This removed the potential of creating intermediate analytic artifacts that could not be shared. All the artifacts were digital and were implicitly shared. This was done to understand how a digital only setup would function, and if it would still be possible to perform sensemaking when pen and paper were unavailable. Future studies and tools should consider including non-digital tools for sensemaking to situate their research in “real world” better. For example, alternative setups might involve mobile location aware devices for collocated sensemaking (Wozniak and Goyal et. al., 2016)

Fourth, despite the iterative nature of the design of the tools, results cannot be directly compared across them. Solo SAVANT involved a single user trying to solve 10 crime cases in 1 hour. Collaborative SAVANT shared 3 and two half crime cases each over an hour, forcing users to spend time and effort sharing information that would have been otherwise directly available. REFLECTIVA task was shorter by 15 minutes, with extra effort and support from the new features. While these experimental modifications were made to continue to ensure that the task was challenging, it made direct result comparison challenging across the iterations. For example, REFLECTIVA standard interface reported slightly lower Clue Recall than SAVANT implicit condition. Readers should be careful about reading results due to the experimental changes.

Finally, while my work has been significantly focussed on crime-solving based domain, similar analytical challenges exist in law, education, epidemiology, health, and space programs. Using intermediate analytic artifacts might still be useful for tasks like disconfirming judgements, enriching qualitative analysis with more interpretations, identifying hidden patterns of spread of diseases, etc. This, open-ended exploratory tasks beyond intelligence analysis can benefit from these approaches to improve social and cognitive processes in collaboration.

Future Work

One particular direction of future work, would be identifying other domains where SAVANT and REFLECTIVA might be useful. One such domain might be qualitative data analysis. Qualitative researchers inductively perform analysis across large bodies of text. They usually parse the same piece of text multiple times until they have exhausted identifying codes and themes in the dataset. They may then try to decipher how themes might be interconnected within each other. This process of qualitative data analysis is an inductive and time-consuming,

but rewarding process, as the researchers might uncover phenomena that could not be discovered using other techniques. SAVANT may be modified to help multiple qualitative researchers code data and share their codes, associated with the same textual pieces. The researchers may then choose to communicate with each other about common codes or use the free form analysis space to identify how multiple themes are connected and help tell the story behind their data. REFLECTIVA may help researchers remain self-aware that they are not focusing on one particular interviewee or even theme. This will help them spread the focus of their attention equally across multiple interviewees and/or themes.

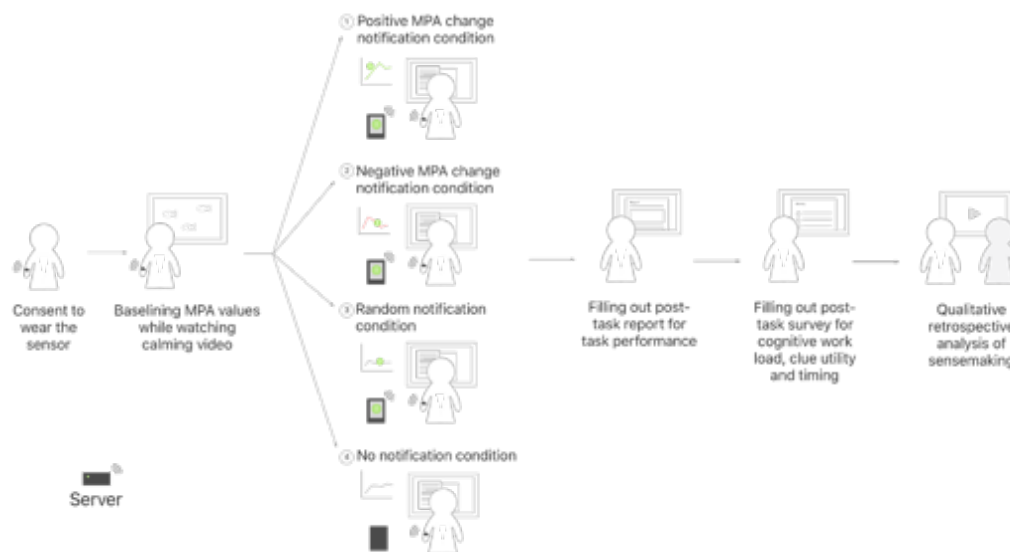


Figure 16 The Experimental Setup. Participant signed a consent form prior to wearing Q Sensors (by Affectiva Inc.). Next, the Q Sensor was synched with a server and EDA values were baselined at the end of watching a calming swimming fish video. Based, on one of the four conditions, user then was explained the task, give the task-set, and was notified appropriately (or not at all in the fourth condition). The sensor was removed at the end of task. The user then filled out a report about serial killer cases and serial killer identity. Next, user filled out a survey about workload, and clue utility and timing. Finally, the experimenter performed the retrospective video analysis of the sensemaking process associated with EDA acceleration, deceleration, and clue timing.

A second direction for future work might be designing to overcome interruptions. As introduced in chapter 1, there was no joint information center to manage interrupting requests by the 24/7 news cycle after the Boston Marathon Bombing, and this led to confusion about what each agency was doing. However, interruptions are necessary to collaboration. While analysts are faced with a constant influx of information, they have to identify the importance of each new piece of information. Previous work has shown that randomly interrupting workers with notifications can decrease performance and that interruptions may have fewer negative consequences if presented at task boundaries (Bailey et al., 2006; Czerwinski et al., 2000).

However, this approach requires developing, modeling and validating the task execution structures prior to manipulating interruptions (Bailey & Iqbal, 2008), and can be difficult for complex analytical tasks (Balakrishnan et al., 2010; Goyal et al., 2016; Goyal et al., 2013; Goyal

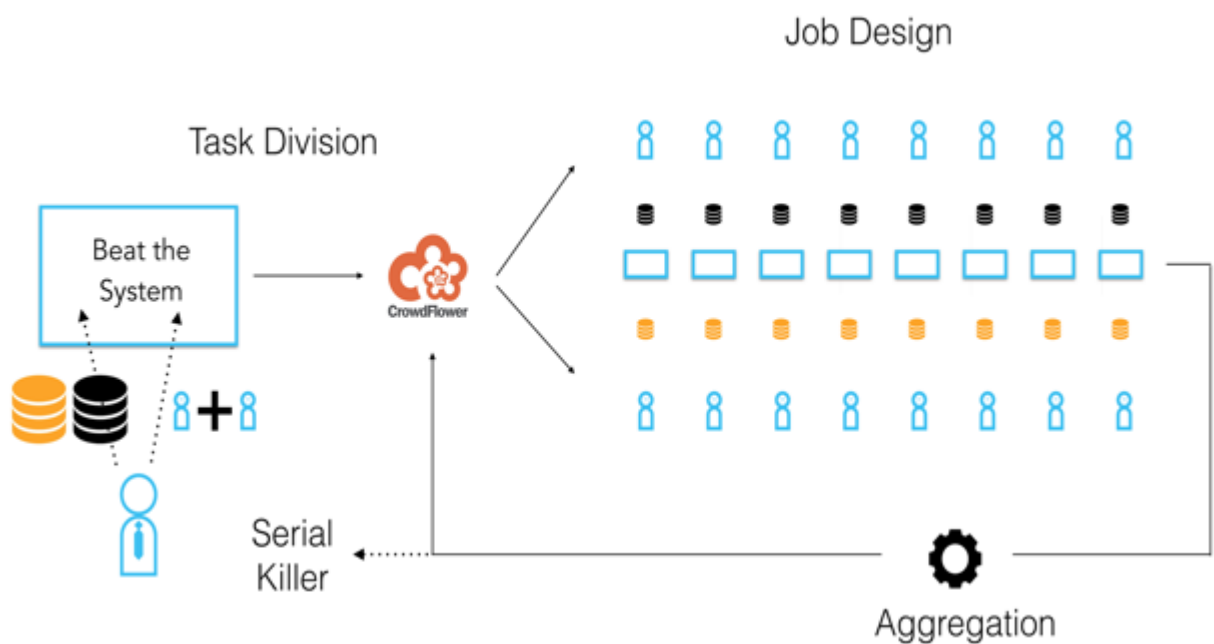


Figure 17: Workflow for Crowdsourcing

et al., 2014). Instead future tools could leverage physiological data from EDA (Electrodermal Activity derived from Galvanic Skin Response) and help identify arousal states (Ax, 1953; Nagai et al., 2004; Sykes et al., 2003, Züger et al. 2015) to infer interruptibility, as shown in a sensemaking study (Figure 16) where interrupting at significant acceleration and deceleration of EDA had a significant impact on task performance (Goyal et al, 2017).

An alternative direction for future work might be enabling non-experts to help experts perform complex analysis. The Boston Marathon Bombing, as mentioned in Chapter 1, involved non-expert crime-solving redditors going online and iteratively trying to piece together the puzzle of who was responsible. While cognitive tunneling probably played a role in their failure to identify the correct criminals, they may also have failed due to a lack of coherent structure and direction. At the same time, there is a lack of sufficient expert resources to handle vast large amounts of crime data. Crowdsourcing platforms are increasingly becoming online platforms tools for enlisting help from workers who are not part of the traditional labor market.

Some researchers have tried crowdsourcing for slightly more complex tasks such as producing reports (Bernstein et al., 2015; Teevan et al., 2016), creating short fiction stories (Kim et al., 2016) and creating travel itineraries (Zhang et al., 2012), or tasks requiring interdependence (Hahn et al., 2016, Salehi et al., 2017). Crowdsourcing crime-solving by non-experts can be even more challenging due to a lack of domain expertise or sufficient domain experts available for training. Future work (Goyal et al, 2015) could leverage workflows that enable non-experts to contribute towards analysis with experts, as shown in Figure 17.

Conclusion

In this dissertation, I presented findings from multiple experiments that iteratively developed the intelligence analysis tools for remote collaboration in small groups. While the previous research work has focused on computational advancement, this dissertation offers a human-centered approach to introduce collaborative sensemaking as a social-cognitive design challenge of effective modular tool design, social information sharing, and overcoming cognitive tunneling

This dissertation suggests leveraging customization to build modular tools, automated implicit sharing to overcome cognitive load, and self-reflective visualizations to expand focus of attention. This dissertation further presents and evaluates designs of multiple interfaces: solo and collaborative SAVANT, and REFLECTIVA to show their effectiveness on sensemaking. This work contributes to design and theory of collaborative sensemaking and opens up new directions for future work.

REFERENCES

- Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when?: The effects of interruption at different moments within task execution. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 271–278.
- Affectiva Q Sensor, 2017 <http://qsensor-support.affectiva.com>
- Andrews, C., Endert, A., & North, C. (2010). Space to think: Large high-resolution displays for sensemaking. *Proceedings of the SIGCHI conference on human factors in computing systems*, 55–64.
- Arnstein, S. (1969). *J. Amer. Inst. Planners*, 35, 216.
- Atzl, C., Meschtscherjakov, A., Vikoler, S., & Tscheligi, M. (2015). Bet4EcoDrive. *International Conference on Persuasive Technology*, 71–82.
- Ax, A. F. (1953). The physiological differentiation between fear and anger in humans. *Psychosomatic medicine*, 15(5), 433–442.
- Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 14(4), 21.
- Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in human behavior*, 22(4), 685–708.
- Bailey, B. P., Adamczyk, P. D., Chang, T. Y., & Chilson, N. A. (2006). A framework for specifying and monitoring user tasks. *Computers in Human Behavior*, 22(4), 709–732.

- Balakrishnan, A. D., Fussell, S. R., & Kiesler, S. (2008). Do visualizations improve synchronous remote collaboration? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1227–1236.
- Balakrishnan, A. D., Fussell, S. R., Kiesler, S., & Kittur, A. (2010). Pitfalls of information access with visualizations in remote collaborative analysis. *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 411–420.
- Bardram, J. E., Hansen, T. R., & Soegaard, M. (2006). AwareMedia: A shared interactive display supporting social, temporal, and spatial awareness in surgery. *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, 109–118.
- Benkler, Y. (2002). Coase's Penguin, or, Linux and "The Nature of the Firm." *Yale Law Journal*, 369–446.
- Bernstein, M. S. (2010). Crowd-powered interfaces. Adjunct proceedings of the 23rd annual ACM symposium on User interface software and technology, 347–350.
- Bernstein, M. S., Little, G., Miller, R. C., Hartmann, B., Ackerman, M. S., Karger, D. R., & Crowell, D. (2015). Soylent: A word processor with a crowd inside. *Communications of the ACM*, 58(8), 85–94.
- Bier, E. A., Card, S. K., & Bodnar, J. W. (2010). Principles and tools for collaborative entity-based intelligence analysis. *IEEE transactions on visualization and computer graphics*, 16(2), 178–191.
- Billman, D., Convertino, G., Pirolli, P., Massar, J. P., & Shrager, J. (2005). Collaborative intelligence analysis with CACHE: Bias reduction and information coverage. *Unpublished manuscript. Palo Alto Research Center, CA.*

- Bostock, M., & Heer, J. (2009). Protovis: A graphical toolkit for visualization. *IEEE transactions on visualization and computer graphics*, 15(6).
- Brush, A. J., Barger, D., Grudin, J., & Gupta, A. (2002). Notification for shared annotation of digital documents. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 89–96.
- Cabrera, A., & Cabrera, E. F. (2002). Knowledge-sharing dilemmas. *Organization studies*, 23(5), 687–710.
- Cai, C. J., Iqbal, S. T., & Teevan, J. (2016). Chain reactions: The impact of order on microtask chains. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3143–3154.
- Carter, D. (2014). Encouraging ambiguous experience: Guides for personal meaning making. *Proceedings of the 2014 companion publication on Designing interactive systems*, 61–64.
- Catlin-Groves, C. (2012). The citizen science landscape: From volunteers to citizen sensors and beyond. *International Journal of Zoology*.
- Chan, J., Dang, S., & Dow, S. P. (2016). Comparing Different Sensemaking Approaches for Large-Scale Ideation. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2717–2728.
- Chang, J. C., Kittur, A., & Hahn, N. (2016). Alloy: Clustering with crowds and computation. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*.
- Chen, D., Hart, J., & Vertegaal, R. (2007). Towards a physiological model of user interruptability. *IFIP Conference on Human-Computer Interaction*, 439–451.

- Chilton, L. B., Little, G., Edge, D., Weld, D. S., & Landay, J. A. (2013). Cascade: Crowdsourcing taxonomy creation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1999–2008.
- Chin, Jr., George, O. A. K., & Wolf, K. E. (2009). Exploring the analytical processes of intelligence analysts. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 11–20.
- Chow, K. K., Harrell, D. F., & Wong Ka Yan. (2015). Designing and analyzing swing compass: A lively interactive system provoking imagination and affect for persuasion. *International Conference on Persuasive Technology*, 107–120.
- Chuah, M. C., & Roth, S. F. (2003). Visualizing common ground. *Information Visualization*, 2003. IV 2003. Proceedings. Seventh International Conference on, 365–372.
- Chung, H., Yang, S., Massjouni, N., Andrews, C., Kanna, R., & North, C. (2010). Vizcept: Supporting synchronous collaboration for constructing visualizations in intelligence analysis. *Visual Analytics Science and Technology (VAST)*, 107–114.
- Cifelli, R., Doesken, N., Kennedy, P., Carey, L. D., Rutledge, S. A., Gimmestad, C., & Depue, T. (2005). The community collaborative rain, hail, and snow network: Informal education for scientists and citizens. *Bulletin of the American Meteorological Society*, 86(8), 1069–1077.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition*, 13, 127–149.

- Consolvo, S., Klasnja, P., McDonald, D. W., & Landay, J. A. (2009). Goal-setting considerations for persuasive technologies that encourage physical activity. *Proceedings of the 4th international Conference on Persuasive Technology*, 8.
- Convertino, G., Billman, D., Pirolli, P., Massar, J. P., & Shrager, J. (2008). The CACHE study: Group effects in computer-supported collaborative analysis. *Computer Supported Cooperative Work (CSCW)*, 17(4), 353–393.
- Convertino, G., Mentis, H. M., Mary Beth Rosson, Slavkovic, A., & Carroll, J. M. (2009). Supporting content and process common ground in computer-supported teamwork. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2339–2348.
- Convertino, G., Mentis, H. M., Mary Beth Rosson, Carroll, J. M., Slavkovic, A., & Ganoe, C. H. (2008). Articulating common ground in cooperative work: Content and process. *proceedings of the SIGCHI conference on human factors in computing systems*, 1637–1646.
- Convertino, G., Mentis, H. M., Slavkovic, A., Mary Beth Rosson, & Carroll, J. M. (2011). Supporting common ground and awareness in emergency management planning: A design research project. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(4), 22.
- Cook, M. B., & Smallman, H. S. (2007). Visual evidence landscapes: Reducing bias in collaborative intelligence analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 51(4), 303–307.

- Cowley, P., Nowell, L., & Scholtz, J. (2005). Glass box: An instrumented infrastructure for supporting human interaction with information. *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*, 296c.
- Crowston, K., & Prestopnik, N. R. (2013). Motivation and data quality in a citizen science game: A design science evaluation. *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, 450–459.
- Culbertson, G., Shen, S., Andersen, E., & Jung, M. F. (2017). Have your Cake and Eat it Too: Foreign Language Learning with a Crowdsourced Video Captioning System. *CSCW*, 286–296.
- Czerwinski, M., Cutrell, E., & Horvitz, E. (2000). Instant messaging: Effects of relevance and timing. *People and computers XIV: Proceedings of HCI*, 2, 71–76.
- Czerwinski, M., Cutrell, E., & Horvitz, E. (2000). Instant messaging and interruption: Influence of task type on performance. *OZCHI 2000 conference proceedings*, 356, 361–367.
- De Boer, P. M., & Bernstein, A. (2017). Efficiently Identifying a Well-Performing Crowd Process for a Given Problem. *CSCW*, 1688–1699.
- de Vries, Roelof AJ, Truong, K. P., Kwint, S., Drossaert, C. H., & Evers, V. (2016). Crowd-Designed Motivation: Motivational messages for exercise adherence based on behavior change theory. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 297–308.
- Dergousoff, K., & Mandryk, R. L. (2015). Mobile gamification for crowdsourcing data collection: Leveraging the freemium model. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1065–1074.

- Deterding, S. (2011). Meaningful play: Getting gamification right. *Google Tech Talk*, 24.
- Ding, Xianghua, Patrick C. Shih, and Ning Gu. "Socially Embedded Work: A Study of Wheelchair Users Performing Online Crowd Work in China." In *CSCW*, pp. 642-654. 2017.
- Dismukes, R. K., Grant E. Young, and Mary Connors. "Cockpit Interruptions and Distractions: An Analysis of ASRS Reports and an Experimental Program." (1998).
- Dittus, Martin, Giovanni Quattrone, and Licia Capra. "Mass Participation During Emergency Response: Event-centric Crowdsourcing in Humanitarian Mapping." In *CSCW*, pp. 1290-1303. 2017.
- Doroudi, Shayan, Ece Kamar, Emma Brunskill, and Eric Horvitz. "Toward a learning science for complex crowdsourcing tasks." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2623-2634. ACM, 2016.
- Dow, S., Kulkarni, A., Bunge, B., Nguyen, T., Klemmer, S., & Hartmann, B. (2011, May). Shepherd the crowd: managing and providing feedback to crowd workers. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 1669-1674). ACM.
- Dwyer, N., & Suthers, D. D. (2005, May). A study of the foundations of artifact-mediated collaboration. In *Proceedings of the 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!*(pp. 135-144). International Society of the Learning Sciences.
- Egger, S. A. (2002). *The killers among us: An examination of serial murder and its investigation*. Upper Saddle River, NJ: Prentice Hall.

- Fisher, K., Counts, S., & Kittur, A. (2012, May). Distributed sensemaking: improving sensemaking by leveraging the efforts of previous users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 247-256). ACM.
- Fogarty, J., Ko, A. J., Aung, H. H., Golden, E., Tang, K. P., & Hudson, S. E. (2005, April). Examining task engagement in sensor-based statistical models of human interruptibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 331-340). ACM.
- Fogg, B. J. (2009, April). Creating persuasive technologies: an eight-step design process. In *Proceedings of the 4th international conference on persuasive technology* (p. 44). ACM.
- Fogg, B. J. (2009, April). A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*(p. 40). ACM.
- Froehlich, J., & Dourish, P. (2004, May). Unifying artifacts and activities in a visual tool for distributed software development teams. In *Proceedings of the 26th International Conference on Software Engineering* (pp. 387-396). IEEE Computer Society.
- Ganoe, C. H., Somervell, J. P., Neale, D. C., Isenhour, P. L., Carroll, J. M., Rosson, M. B., & McCrickard, D. S. (2003, November). Classroom BRIDGE: using collaborative public and desktop timelines to support activity awareness. In *Proceedings of the 16th annual ACM symposium on User interface software and technology* (pp. 21-30). ACM.
- General Dynamics (2011). Co-Motion. <http://www.gdc4s.com/content/detail.cfm?item=32341561-76f9-40f8-8ad5-0f0d66dd240e>

- Gergle, D., Kraut, R. E., & Fussell, S. R. (2004). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of language and social psychology*, 23(4), 491-517.
- Gottlieb, S., Arenberg, S. I., & Singh, R. (1994). *Crime analysis: From first report to final arrest*. Montclair, CA: Alpha Publishing.
- Goyal, N., & Fussell, S. R. (2015). Designing for Collaborative Sensemaking: Leveraging Human Cognition For Complex Tasks. *arXiv preprint arXiv:1511.05737*.
- Goyal, N., & Fussell, S. R. (2016, February). Effects of Sensemaking Translucence on Distributed Collaborative Analysis. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 288-302). ACM.
- Goyal, N., Leshed, G., & Fussell, S. R. (2013, April). Effects of visualization and note-taking on sensemaking and analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2721-2724). ACM.
- Goyal, N., Leshed, G., & Fussell, S. R. (2013, February). Leveraging partner's insights for distributed collaborative sensemaking. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*(pp. 15-18). ACM.
- Goyal, N., Leshed, G., Cosley, D., & Fussell, S. R. (2014, April). Effects of implicit sharing in collaborative analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 129-138). ACM.
- Goyal, N., & Fussell, S. R. (2017). Intelligent Interruption Management using Electro Dermal Activity based Physiological Sensor on Collaborative Sensemaking. In *Proceedings of the ACM IMWUT 3.1, 2017*

- Gutwin, C., & Greenberg, S. (1998, November). Design for individuals, design for groups: tradeoffs between power and workspace awareness. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work* (pp. 207-216). ACM.
- Hahn, N., Chang, J., Kim, J. E., & Kittur, A. (2016, May). The Knowledge Accelerator: Big picture thinking in small pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2258-2270). ACM.
- Hailpern, J., Hinterbichler, E., Leppert, C., Cook, D., & Bailey, B. P. (2007, June). TEAM STORM: demonstrating an interaction model for working with multiple ideas during creative group work. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition* (pp. 193-202). ACM.
- Harjumaa, M., Segerståhl, K., & Oinas-Kukkonen, H. (2009, April). Understanding persuasive software functionality in practice: a field trial of polar FT60. In *proceedings of the 4th international conference on persuasive technology* (p. 2). ACM.
- Hart, S. G., & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale. *Human mental workload*, 139-183.
- Hata, K., Krishna, R., Fei-Fei, L., & Bernstein, M. S. (2016). A Glimpse Far into the Future: Understanding Long-term Crowd Worker Accuracy. *arXiv preprint arXiv:1609.04855*.
- Hatcher, E., & Gospodnetic, O. (2004). Lucene in action.
- Hayne, S. C., Troup, L. J., & Mccomb, S. A. (2011). “Where’s Farah?”: Knowledge silos and information fusion by distributed collaborating teams. *Information Systems Frontiers*, 13(1), 89-100.

- Heer, J., & Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information visualization*, 7(1), 49-62.
- Heer, J., Viégas, F. B., & Wattenberg, M. (2009). Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Communications of the ACM*, 52(1), 87-97.
- Heimerl, K., Gawalt, B., Chen, K., Parikh, T., & Hartmann, B. (2012, May). CommunitySourcing: engaging local crowds to perform expert work via physical kiosks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1539-1548). ACM.
- Heuer, R. J. (1999). Psychology of intelligence analysis. Lulu. com.
- Hincapié Ramos, J. D., Tabard, A., & Bardram, J. E. (2011, May). Gridorbit: an infrastructure awareness system for increasing contribution in volunteer computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1899-1908). ACM.
- Hoffman, G., Birnbaum, G. E., Vanunu, K., Sass, O., & Reis, H. T. (2014, March). Robot responsiveness to human disclosure affects social impression and appeal. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 1-8). ACM.
- Horvitz, E., Jacobs, A., & Hovel, D. (1999, July). Attention-sensitive alerting. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence* (pp. 305-313). Morgan Kaufmann Publishers Inc.
- Hu, M., Yang, H., Zhou, M., Gou, L., Li, Y., & Haber, E. (2013, September). OpinionBlocks: a crowd-powered, self-improving interactive visual analytic system for understanding

- opinion text. In *14th International Conference on Human-Computer Interaction (INTERACT)* (No. Part II, pp. 116-134). Springer.
- i2 (2011). Analyst's Notebook. www.i2group.com/us
- Ikeda, K., & Bernstein, M. S. (2016, May). Pay It Backward: Per-Task Payments on Crowdsourcing Platforms Reduce Productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4111-4121). ACM.
- Iqbal, S. T., & Bailey, B. P. (2008, April). Effects of intelligent notification management on users and their tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 93-102). ACM.
- Iqbal, S. T., & Bailey, B. P. (2006, April). Leveraging characteristics of task structure to predict the cost of interruption. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 741-750). ACM.
- Isenberg, P., Fisher, D., Paul, S. A., Morris, M. R., Inkpen, K., & Czerwinski, M. (2012). Co-located collaborative visual analytics around a tabletop display. *IEEE Transactions on visualization and Computer Graphics*, 18(5), 689-702.
- Jackson, T., Dawson, R., & Wilson, D. (2001). The cost of email interruption. *Journal of Systems and Information Technology*, 5(1), 81-92.
- Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning?. *Computers & Education*, 49(4), 1037-1065.

- Jarvenpaa, S. L., & Staples, D. S. (2000). The use of collaborative electronic media for information sharing: an exploratory study of determinants. *The Journal of Strategic Information Systems*, 9(2), 129-154.
- Jensen, E. (2009). Sensemaking in military planning: a methodological study of command teams. *Cognition, Technology & Work*, 11(2), 103-118.
- Johnston, R. (2005). *Analytic culture in the US intelligence community: An ethnographic study*. CENTRAL INTELLIGENCE AGENCY WASHINGTON DC CENTER FOR STUDY OF INTELLIGENCE.
- Jung, J. H., Schneider, C., & Valacich, J. (2010). Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management science*, 56(4), 724-742.
- Kam, M., Wang, J., Iles, A., Tse, E., Chiu, J., Glaser, D., ... & Canny, J. (2005, April). Livenotes: a system for cooperative and augmented note-taking in lectures. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 531-540). ACM.
- Kandappu, T., Misra, A., & DARATAN, R. T. (2017). Collaboration trumps homophily in urban mobile crowd-sourcing.
- Kang, R., Kane, A., & Kiesler, S. (2014, February). Teammate inaccuracy blindness: when information sharing tools hinder collaborative analysis. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 797-806). ACM.

- Kang, R., & Kiesler, S. (2012, February). Do collaborators' annotations help or hurt asynchronous analysis. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion* (pp. 123-126). ACM.
- Kang, Y. A., Gorg, C., & Stasko, J. (2011). How can visual analytics assist investigative analysis? Design implications from an evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 17(5), 570-583.
- Kaufman, G., Flanagan, M., & Punjasthitkul, S. (2016, May). Investigating the impact of 'emphasis frames' and social loafing on player motivation and performance in a crowdsourcing game. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4122-4128). ACM.
- Keel, P. E. (2007). EWall: A visual analytics environment for collaborative sense-making. *Information Visualization*, 6(1), 48-63.
- Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., Popović, Z., & Baker, D. (2011). Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences*, 108(47), 18949-18953.
- Kim, J., Stermann, S., Cohen, A. A. B., & Bernstein, M. S. (2016). Mechanical novel: Crowdsourcing complex work through reflection and revision. *arXiv preprint arXiv:1611.02682*.
- Kim, S., & Paulos, E. (2010, April). InAir: sharing indoor air quality measurements and visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1861-1870). ACM.

- Kim, S., Robson, C., Zimmerman, T., Pierce, J., & Haber, E. M. (2011, May). Creek watch: pairing usefulness and usability for successful citizen science. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2125-2134). ACM.
- Kittur, A., Smus, B., Khamkar, S., & Kraut, R. E. (2011, October). Crowdforge: Crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 43-52). ACM.
- Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground and coordination in joint activity. *Organizational simulation*, 53, 139-184.
- Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent systems*, 21(5), 88-92.
- Kriglstein, S., Wallner, G., & Pohl, M. (2014, April). A user study of different gameplay visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 361-370). ACM.
- Krishna, R. A., Hata, K., Chen, S., Kravitz, J., Shamma, D. A., Fei-Fei, L., & Bernstein, M. S. (2016, May). Embracing error to enable rapid crowdsourcing. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 3167-3179). ACM.
- Kulkarni, A., Can, M., & Hartmann, B. (2012, February). Collaboratively crowdsourcing workflows with turkomatic. In *Proceedings of the acm 2012 conference on computer supported cooperative work* (pp. 1003-1012). ACM.
- Lakhani, K. R., & Wolf, R. G. (2003). Why hackers do what they do: Understanding motivation and effort in free/open source software projects.

- Lasecki, W. S., Wesley, R., Nichols, J., Kulkarni, A., Allen, J. F., & Bigham, J. P. (2013, October). Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM symposium on User interface software and technology* (pp. 151-162). ACM.
- Latorella, K. A. (1998, October). Effects of modality on interrupted flight deck performance: Implications for data link. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 42, No. 1, pp. 87-91). Sage CA: Los Angeles, CA: SAGE Publications.
- Laurence, S. X., Attfield, S., & Fields, B. (2016, July). Common ground in collaborative intelligence analysis: an empirical study. In *Proceedings of the 30th International BCS Human Computer Interaction Conference: Fusion!* (p. 30). BCS Learning & Development Ltd.
- Law, E., Gajos, K. Z., Wiggins, A., Gray, M. L., & Williams, A. C. (2017, February). Crowdsourcing as a Tool for Research: Implications of Uncertainty. In *CSCW* (pp. 1544-1561).
- Lee, J., & Rao, H. R. (2007, May). Exploring the causes and effects of inter-agency information sharing systems adoption in the anti/counter-terrorism and disaster management domains. In *Proceedings of the 8th annual international conference on Digital government research: bridging disciplines & domains* (pp. 155-163). Digital Government Society of North America.
- Lee, J. D., Hoffman, J. D., & Hayes, E. (2004, April). Collision warning design to mitigate driver distraction. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems* (pp. 65-72). ACM.

- Little, G., Chilton, L. B., Goldman, M., & Miller, R. C. (2010, October). TurkIt: human computation algorithms on mechanical Turk. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology* (pp. 57-66). ACM.
- Lockton, D., Harrison, D., Holley, T., & Stanton, N. A. (2009, April). Influencing interaction: development of the design with intent method. In *proceedings of the 4th international conference on persuasive technology* (p. 5). ACM.
- Luther, K., Counts, S., Stecher, K. B., Hoff, A., & Johns, P. (2009, April). Pathfinder: an online collaboration environment for citizen scientists. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 239-248). ACM.
- Maisonneuve, N., Stevens, M., Niessen, M. E., Hanappe, P., & Steels, L. (2009, May). Citizen noise pollution monitoring. In *Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government* (pp. 96-103). Digital Government Society of North America.
- Malone, T. W., Laubacher, R., & Dellarocas, C. (2009). Harnessing crowds: Mapping the genome of collective intelligence.
- Mamykina, L., Smyth, T. N., Dimond, J. P., & Gajos, K. Z. (2016, May). Learning from the crowd: Observational learning in crowdsourcing communities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2635-2644). ACM.
- Mankowski, T. A., Slater, S. J., & Slater, T. F. (2011). An interpretive study of meanings citizen scientists make when participating in Galaxy Zoo. *Contemporary Issues in Education Research*, 4(4), 25.

- Mark, G., Carpenter, K., & Kobsa, A. (2003, July). A model of synchronous collaborative information visualization. In *Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on* (pp. 373-381). IEEE.
- Massung, E., Coyle, D., Cater, K. F., Jay, M., & Preist, C. (2013, April). Using crowdsourcing to support pro-environmental community activism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 371-380). ACM.
- Mathan, S., Whitlow, S., Dorneich, M., Ververs, P., & Davis, G. (2007, October). Neurophysiological estimation of interruptibility: Demonstrating feasibility in a field context. In *Proceedings of the 4th International Conference of the Augmented Cognition Society* (pp. 51-58).
- McCarthy, J. C., Miles, V. C., & Monk, A. F. (1991, April). An experimental study of common ground in text-based communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 209-215). ACM.
- McCrickard, D. S., Catrambone, R., Chewar, C. M., & Stasko, J. T. (2003). Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems. *International Journal of Human-Computer Studies*, 58(5), 547-582.
- McFarlane, D. C., & Latorella, K. A. (2002). The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction*, 17(1), 1-61.
- McInnis, B., Cosley, D., Nam, C., & Leshed, G. (2016, May). Taking a HIT: Designing around rejection, mistrust, risk, and workers' experiences in Amazon Mechanical Turk. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 2271-2282). ACM.

- Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013, April). Disassembling gamification: the effects of points and meaning on user motivation and performance. In *CHI'13 extended abstracts on human factors in computing systems* (pp. 1137-1142). ACM.
- Mentis, H. M., Bach, P. M., Hoffman, B., Rosson, M. B., & Carroll, J. M. (2009, April). Development of decision rationale in complex group decision making. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1341-1350). ACM.
- Merritt, David, Jasmine Jones, Mark S. Ackerman, and Walter S. Lasecki. "Kurator: Using The Crowd to Help Families With Personal Curation Tasks." In *CSCW*, pp. 1835-1849. 2017.
- Myers, B. A. (1985, April). The importance of percent-done progress indicators for computer-human interfaces. In *ACM SIGCHI Bulletin* (Vol. 16, No. 4, pp. 11-17). ACM.
- Nagai, Y., Goldstein, L. H., Critchley, H. D., & Fenwick, P. B. (2004). Influence of sympathetic autonomic arousal on cortical arousal: implications for a therapeutic behavioural intervention in epilepsy. *Epilepsy research*, 58(2), 185-193.
- National Audubon Society (2010), Christmas Bird Count. birds.audubon.org/history-christmas-bird-count.
- National Commission on Terrorist Attacks upon the United States, 2004. The 9/11 Commission Report: Final report of the national commission on terrorist attacks upon the united states, Norton, NY.
- Newell, E., & Ruths, D. (2016, May). How one microtask affects another. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3155-3166). ACM.

- New York Times, 2013: <https://www.nytimes.com/topic/subject/boston-marathon-bombings>
- New York Times, 2017: <https://www.nytimes.com/2017/06/06/world/europe/london-assailants-terrorism-warning-signs-fbi.html>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175.
- Nobarany, S., Haraty, M., & Fisher, B. (2012, February). Facilitating the reuse process in distributed collaboration: a distributed cognition approach. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 1223-1232). ACM.
- Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012, November). Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (pp. 420-423). ACM.
- Nov, O., Anderson, D., & Arazy, O. (2010, April). Volunteer computing: a model of the factors determining contribution to community-based scientific research. In *Proceedings of the 19th international conference on World wide web* (pp. 741-750). ACM.
- Nov, O., Arazy, O., & Anderson, D. (2011, July). Technology-Mediated Citizen Science Participation: A Motivational Model. In *ICWSM*.
- Nov, O. (2007). What motivates wikipedians?. *Communications of the ACM*, 50(11), 60-64.
- Okoshi, T., Ramos, J., Nozaki, H., Nakazawa, J., Dey, A. K., & Tokuda, H. (2015, September). Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile environments. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 475-486). ACM.
- Palantir Technology (2011). www.palantirtech.com

- Park, C. H., Son, K., Lee, J. H., & Bae, S. H. (2013, February). Crowd vs. crowd: large-scale cooperative design through open team competition. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 1275-1284). ACM.
- Paul, S. A., & Reddy, M. C. (2010, February). Understanding together: sensemaking in collaborative information seeking. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 321-330). ACM.
- Pejovic, V., & Musolesi, M. (2014, September). InterruptMe: designing intelligent prompting mechanisms for pervasive applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 897-908). ACM.
- Perer, A., & Shneiderman, B. (2008, January). Systematic yet flexible discovery: guiding domain experts through exploratory data analysis. In *Proceedings of the 13th international conference on Intelligent user interfaces*(pp. 109-118). ACM.
- Petersen, R. R., & Wiil, U. K. (2011, September). Crimefighter investigator: A novel tool for criminal network investigation. In *Intelligence and Security Informatics Conference (EISIC), 2011 European* (pp. 197-202). IEEE.
- Pickton Report: http://www.ag.gov.bc.ca/public_inquiries/docs/Forsaken-ES.pdf. Pickton Trials: <http://www.cbc.ca/news/canada/british-columbia/story/2007/01/22/pickton-trial.html>
- Pioch, N. J., & Everett, J. O. (2006, November). POLESTAR: collaborative knowledge management and sensemaking tools for intelligence analysts. In *Proceedings of the 15th ACM international conference on Information and knowledge management* (pp. 513-521). ACM.

- Pirolli, P., & Card, S. (2005, May). Sensemaking processes of intelligence analysts and possible leverage points as identified through cognitive task analysis. In *Proceedings of the 2005 International Conference on Intelligence Analysis, McLean, Virginia* (Vol. 6).
- Pirolli, P., & Card, S. (1995, May). Information foraging in information access environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 51-58). ACM Press/Addison-Wesley Publishing Co.
- Pirolli, P., & Card, S. (1999). Information foraging. *Psychological review*, 106(4), 643.
- Pirolli, P., & Card, S. (2005, May). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (Vol. 5, pp. 2-4).
- Police Chief magazine*, October 2009
http://www.policechiefmagazine.org/magazine/index.cfm?fuseaction=display_arch&article_id=1922&issue_id=102009
- Prante, T., Magerkurth, C., & Streitz, N. (2002, November). Developing CSCW tools for idea finding-: empirical results and implications for design. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work* (pp. 106-115). ACM.
- Preist, C., Massung, E., & Coyle, D. (2014, February). Competing or aiming to be average?: normification as a means of engaging digital volunteers. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 1222-1233). ACM.
- Quake Catcher Network: <http://qcn.stanford.edu>

- Quattrone, G., Dittus, M., & Capra, L. (2017, February). Work Always in Progress: Analysing Maintenance Practices in Spatial Crowd-sourced Datasets. In *CSCW* (pp. 1876-1889).
- Raddick, J., Lintott, C. J., Schawinski, K., Thomas, D., Nichol, R. C., Andreescu, D., ... & Szalay, A. S. (2007, December). Galaxy Zoo: an experiment in public science participation. In *Bulletin of the American Astronomical Society* (Vol. 39, p. 892).
- Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., ... & Vandenberg, J. (2009). Galaxy zoo: Exploring the motivations of citizen science volunteers. *arXiv preprint arXiv:0909.2925*.
- Reddy, M., & Dourish, P. (2002, November). A finger on the pulse: temporal rhythms and information seeking in medical work. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work* (pp. 344-353). ACM.
- Reed, J., Raddick, M. J., Lardner, A., & Carney, K. (2013, January). An exploratory factor analysis of motivations for participating in Zooniverse, a collection of virtual citizen science projects. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 610-619). IEEE.
- Reeves, N., Tinati, R., Zerr, S., Simperl, E., & Van Kleek, M. (2017). From crowd to community: a survey of online community features in citizen science projects.
- Retelny, D., Robaszkiewicz, S., To, A., Lasecki, W. S., Patel, J., Rahmati, N., ... & Bernstein, M. S. (2014, October). Expert crowdsourcing with flash teams. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (pp. 75-85). ACM.
- Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., ... & Jacobs, D. (2012, February). Dynamic changes in motivation in collaborative citizen-science projects.

- In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (pp. 217-226). ACM.
- Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993, May). The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 269-276). ACM.
- Salehi, N., McCabe, A., Valentine, M., & Bernstein, M. (2016). Huddler: Convening Stable and Familiar Crowd Teams Despite Unpredictable Availability. *arXiv preprint arXiv:1610.08216*.
- Salehi, N., Teevan, J., Iqbal, S. T., & Kamar, E. (2017, February). Communicating Context to the Crowd for Complex Writing Tasks. In *CSCW*(pp. 1890-1901).
- Santos, R. B. (2016). *Crime analysis with crime mapping*. Sage publications.
- Schneider, C., & von Briel, F. (2013). Crowdsourcing large-scale ecological monitoring: identifying design principles to motivate contributors. In *Building Sustainable Information Systems* (pp. 509-518). Springer, Boston, MA.
- Scupelli, P., Fussell, S. R., Kiesler, S., Quinones, P., & Kusbit, G. (2007, January). Juggling Work Among Multiple Projects and Partner. In *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on* (pp. 77-77). IEEE.
- Sheppard, S. A., & Terveen, L. (2011, October). Quality is a verb: the operationalization of data quality in a citizen science community. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (pp. 29-38). ACM.
- Shih, P. C., Nguyen, D. H., Hirano, S. H., Redmiles, D. F., & Hayes, G. R. (2009, May). GroupMind: supporting idea generation through a collaborative mind-mapping tool.

- In *Proceedings of the ACM 2009 international conference on Supporting group work* (pp. 139-148). ACM.
- Shrinivasan, Y. B., & van Wijk, J. J. (2008, April). Supporting the analytical reasoning process in information visualization. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1237-1246). ACM.
- Shrinivasan, Y. B., & van Wijk, J. (2009). Supporting exploration awareness in information visualization. *IEEE Computer Graphics and Applications*, 29(5), 34-43.
- Sierhuis, M., & Shum, S. B. (2008). Human-agent knowledge cartography for e-science: NASA field trials at the Mars Desert Research Station. In *Knowledge Cartography* (pp. 287-305). Springer London.
- Simon, H. A. (1957). Models of man; social and rational.
- Song, D., & Goldberg, K. (2007). Networked robotic cameras for collaborative observation of natural environments. *Robotics Research*, 510-519.
- Soo Yi, J., Melton, R., Stasko, J., & Jacko, J. A. (2005). Dust & magnet: multivariate information visualization using a magnet metaphor. *Information visualization*, 4(4), 239-256.
- Speier, C., Valacich, J. S., & Vessey, I. (1999). The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, 30(2), 337-360.
- Stanton, N. A. (Ed.). (1994). *Human factors in alarm design*. CRC Press.
- Stasko, J., Görg, C., & Liu, Z. (2008). Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2), 118-132.

- Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of personality and social psychology*, 48(6), 1467.
- Stoyanov, V., Cardie, C., & Wiebe, J. (2005, October). Multi-perspective question answering using the OpQA corpus. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing* (pp. 923-930). Association for Computational Linguistics.
- Suzuki, R., Salehi, N., Lam, M. S., Marroquin, J. C., & Bernstein, M. S. (2016, May). Atelier: Repurposing expert crowdsourcing tasks as micro-internships. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2645-2656). ACM.
- Sykes, J., & Brown, S. (2003, April). Affective gaming: measuring emotion through the gamepad. In *CHI'03 extended abstracts on Human factors in computing systems* (pp. 732-733). ACM.
- Tang, A., Tory, M., Po, B., Neumann, P., & Carpendale, S. (2006, April). Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1181-1190). ACM.
- Taylor, B., Dey, A. K., Siewiorek, D., & Smailagic, A. (2016, May). Using Crowd Sourcing to Measure the Effects of System Response Delays on User Engagement. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4413-4422). ACM.

- Teevan, J., Iqbal, S. T., & Von Veh, C. (2016, May). Supporting collaborative writing with microtasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2657-2668). ACM.
- Tomasic, A., Zimmerman, J., Steinfeld, A., & Huang, Y. (2014, February). Motivating contribution in a participatory sensing system via quid-pro-quo. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 979-988). ACM.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.
- Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making* (pp. 141-162). Springer Netherlands.
- Uchoa, A. P., Esteves, M. G. P., & de Souza, J. M. (2013, June). Mix4Crowds-Toward a framework to design crowd collaboration with science. In *Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on* (pp. 61-66). IEEE.
- Urry, H. L., van Reekum, C. M., Johnstone, T., & Davidson, R. J. (2009). Individual differences in some (but not all) medial prefrontal regions reflect cognitive demand while regulating unpleasant emotion. *Neuroimage*, 47(3), 852-863.
- Valdes, C., Ferreira, M., Feng, T., Wang, H., Tempel, K., Liu, S., & Shaer, O. (2012, November). A collaborative environment for engaging novices in scientific inquiry. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces* (pp. 109-118). ACM.

- Valentine, M. A., Retelny, D., To, A., Rahmati, N., Doshi, T., & Bernstein, M. S. (2017, May). Flash Organizations: Crowdsourcing Complex Work by Structuring Crowds As Organizations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3523-3537). ACM.
- Vogt, K., Bradel, L., Andrews, C., North, C., Endert, A., & Hutchings, D. (2011). Co-located collaborative sensemaking on a large high-resolution display with multiple input devices. *Human-Computer Interaction–INTERACT 2011*, 589-604.
- Waltz, E. (2003). Knowledge management in the intelligence enterprise. Artech House.
- Weaver, C. (2007, July). Is coordination a means to collaboration?. In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on* (pp. 80-84). IEEE.
- Weick, K. E. (1993). The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative science quarterly*, 628-652.
- Whiting, M. E., Gamage, D., Gaikwad, S. S., Gilbee, A., Goyal, S., Ballav, A., ... & Sarma, T. S. (2016). Crowd guilds: Worker-led reputation and feedback on crowdsourcing platforms. *arXiv preprint arXiv:1611.01572*.
- Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2), 165-210.
- Wiggins, A., & Crowston, K. (2011, January). From conservation to crowdsourcing: A typology of citizen science. In *System Sciences (HICSS), 2011 44th Hawaii international conference on* (pp. 1-10). IEEE.

- Willett, W., Heer, J., & Agrawala, M. (2012, May). Strategies for crowdsourcing social data analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 227-236). ACM.
- Willett, W., Heer, J., Hellerstein, J., & Agrawala, M. (2011, May). CommentSpace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 3131-3140). ACM.
- Wong, P. C., Rose, S. J., Chin Jr, G., Frincke, D. A., May, R., Posse, C., ... & Thomas, J. (2006). Walking the path: a new journey to explore and discover through visual analytics. *Information Visualization*, 5(4), 237-249.
- Wozniak, P., Goyal, N., Kucharski, P., Lischke, L., Mayer, S. and Fjeld, M., 2016, May. RAMPARTS: Supporting Sensemaking with Spatially-Aware Mobile Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2447-2460). ACM.
- Wright, W., Schroh, D., Proulx, P., Skaburskis, A., & Cort, B. (2006, April). The Sandbox for analysis: concepts and methods. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 801-810). ACM.
- Wu, Y. W., & Bailey, B. P. (2016, May). Novices Who Focused or Experts Who Didn't?. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4086-4097). ACM.
- Wunsch, M., Stibe, A., Millonig, A., Seer, S., Dai, C., Schechtner, K., & Chin, R. C. (2015, June). What makes you bike? Exploring persuasive strategies to encourage low-energy

- mobility. In *International Conference on Persuasive Technology* (pp. 53-64). Springer, Cham.
- Xie, B., & Salvendy, G. (2000). Prediction of mental workload in single and multiple tasks environments. *International journal of cognitive ergonomics*, 4(3), 213-242.
- Xu, J. J., & Chen, H. (2005). CrimeNet explorer: a framework for criminal network knowledge discovery. *ACM Transactions on Information Systems (TOIS)*, 23(2), 201-226.
- Yu, L., André, P., Kittur, A., & Kraut, R. (2014, February). A comparison of social, learning, and financial strategies on crowd engagement and output quality. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 967-978). ACM.
- Zhang, H., Law, E., Miller, R., Gajos, K., Parkes, D., & Horvitz, E. (2012, May). Human computation tasks with global constraints. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 217-226). ACM.
- Zijlstra, F. R., Roe, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology*, 72(2), 163-185.
- Zilouchian Moghaddam, R., Bailey, B., & Fu, W. T. (2012, May). Consensus building in open source user interface design discussions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1491-1500). ACM.
- Züger, M., & Fritz, T. (2015, April). Interruptibility of software developers and its prediction using psycho-physiological sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2981-2990). ACM.

APPENDIX A.1

* 1. Subject Number

This survey will ask you some questions about your experience of participating in the Detective Experiment. It should take approximately 10 minutes to complete.

The first part of the survey contains a few brief word problems. Please record your answer to the following 3 questions in the text boxes provided.

* 2. A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?

* 3. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

* 4. In a lake there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take the patch to cover half of the lake?

The second part of this survey is to ask you about how much work you thought this experiment required.

* 5. Please select a number from 1 (LOW) to 7 (HIGH) to answer the following questions.

	(Low) 1	2	3	4	5	6	(High) 7
MENTAL DEMAND: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
TEMPORAL DEMAND: How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
EFFORT: How hard did you have to work (mentally and physically) to accomplish your level of performance?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERFORMANCE: How successful do you think you were in accomplishing the goals of the task set by the experimenter?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
FRUSTRATION: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Here we would like to know the kind of Savant system you used during the experiment.

* 6. Was your Analysis Space shared ? (Could you see the stickies from your partner in your analysis space)

- ☐ Yes
☐ No
☐ Not Sure

Here, we would like to ask you a few details about your usage of the Analysis Space

* 7. Select which of the following best characterized your usage of the Hypothesis Window?

	Never	Rarely	Sometimes	Frequently	Very frequently	N/A
I created Hypothesis in the Analysis Space	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I confirmed Hypothesis in the Analysis Space	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I disconfirmed Hypothesis in the Analysis Space	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please let us know about other ways that you used the Hypothesis Window in the box below

* 8. Select which of the following best characterized your usage of the Visualization?

	Never	Rarely	Sometimes	Frequently	Very frequently	N/A
I paid attention to new Avatars in the Analysis Space	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I used Avatars to consider potential suspects	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I used Avatars to rule out potential suspects	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please let us know about other ways that you used the Visualization in the box below

* 9. How, When, and Why did you use/NOT use the Avatars and Visualization at the bottom of the Analysis Space to aid your sensemaking?

* 10. How, When, and Why did you use/NOT use the Hypothesis Window at the bottom of the Analysis Space to aid your sensemaking?

* 11. Please reflect upon your experience with Hypothesis Window in Analysis Space when answering the following questions.

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
The Hypothesis Window made it easy to discuss the cases with my partner.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The Hypothesis Window helped me understand what my partner was thinking .	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The Hypothesis Window made me feel cognitively closer to my partner.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The Hypothesis Window made me feel emotionally closer to my partner.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* 12. Please reflect upon your experience with Visualization in Analysis Space when answering the following questions.

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
The Visualization helped me track my progress	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The Visualization helped me focus on one person	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The Visualization helped me notice persons I did not pay enough attention to	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The visualization did not help me narrow onto one person	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

We are interested in understanding what parts of our tool you used and how helpful to your task you found them.

* 13. Rate how helpful each of the following interface element was when working on solving the cases?

Please rate each tool along the scale provided.

Tools in Document Space (left screen)

	1 (Useless)	2 (Somewhat Useless)	3 (Neither Useless nor Useful)	4 (Somewhat Useful)	5 (Useful)
Network-Diagram (at the top-right for the connections between case documents)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Map (at the bottom-right for stickies and location pointers)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Timeline (at the bottom-left for stickies between time-periods)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stickies (ONLY in the Document space, in the center of the screen)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please explain each choice made above.

* 14. Tools in Analysis Space (right screen)

	Useless	Somewhat Useless	Neither Useless nor Useful	Somewhat Useful	Useful
Chat (at the bottom-left to chat with the partner)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stickies (ONLY in the Analysis space)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hypothesis Window (ONLY in the Analysis space, in the center of the screen)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Visualization (ONLY in the Analysis space, at bottom of the screen)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please explain each choice made above.

Here we would like to hear from you about the entire SAVANT system.

- * 15. If you had a chance to remove one functionality/element/facet/tool of SAVANT (the entire system including the left and right screens), what would you like to remove?

- * 16. If you had a chance to add a functionality/element/facet/tool of SAVANT (the entire system including the left and right screens), what would you like to add?

Now we will ask you some questions about the documents in your cases. We are interested in what kinds of information stood out.

Please answer the following questions regarding the Moeller homicide, even if you were not assigned to this case.

Case description:

Madeline Moeller, an employee at Allegheny General Hospital, was found dead by her husband on the front porch of their residence.

- * 17. A neighbor reported seeing the following on the day that Madeline was murdered:

- ☐ a man exiting the bus carrying a toolbox.
- ☐ a man hanging around outside Madeline's house before she got home.
- ☐ Madeline arguing loudly with her husband as they left the house for work.
- ☐ Madeline catching a ride to work with a female who stopped at her house.
- ☐ I do not know.

* 18. A list of bus riders in the Moeller case

- ☐ did not provide any useful information to solve the serial killer case.
- ☐ included the name of Wayne Millican, who was unemployed at the time.
- ☐ indicated that a business partner of Moeller's husband had been on the bus that day.
- ☐ indicated that a known sex offender, James Stuart Parsley, had been on the bus that day.
- ☐ I do not know.

Please answer the following questions regarding the Fisk homicide, even if you were not assigned to this case.

Case description:

Vicki Fisk, a student at ITT Tech, was found dead near her home by her roommate.

* 19. Vicki Fisk

- ☐ worked as a bartender in Shadyside.
- ☐ had just come to Pittsburgh from Chicago and was unemployed.
- ☐ was murdered between 10 and 10:30 PM.
- ☐ had called her boyfriend on her cell phone from the bus shortly before she was murdered.
- ☐ I do not know.

* 20. Julia Tutt

- ☐ lived with Vicki Fisk near a bus stop.
- ☐ lived with her parents but was hoping to get her own apartment soon.
- ☐ didn't like living in Pittsburgh.
- ☐ complained to friends that she didn't feel safe in Highland Park.
- ☐ I do not know.

Please answer the following questions regarding the Stokes homicide, even if you were not assigned to this case.

Case description:

William T.K. Stokes was shot twice while standing on a sidewalk. Two neighbors witnessed the crime.

* 21. T.K. Stokes

- ☐ had served time for burglary and had recently gotten out of prison.
- ☐ was about to start attending college before he was murdered.
- ☐ had been in a gang but had been trying to get out of it.
- ☐ was known to volunteer at the Boys and Girls Club.
- ☐ I do not know.

* 22. The name of T.K. Stokes's girlfriend was

- ☐ Brittany West
- ☐ Tanya Harmon
- ☐ Janice Truett
- ☐ Sherese McDonald
- ☐ I do not know.

Please answer the following questions regarding the Waring homicide, even if you were not assigned to this case.

Case description:

Myra Waring, a Mercy Hospital employee, was found by her sister with multiple wounds to the head by a blunt instrument.

* 23. Myra Waring

- ☐ was killed by her ex-husband in a fit of rage.
- ☐ was killed around 4 PM, shortly after she got off her shift at Mercy Hospital.
- ☐ was killed on the way back from a jog in her neighborhood one morning.
- ☐ was missing for several days before her body was found.
- ☐ I do not know.

* 24. Gregory Waring's alibi for his ex-wife's murder was:

- ☐ he was at work.
- ☐ he was out of town.
- ☐ he was at a dentist appointment.
- ☐ he didn't have an alibi.
- ☐ I do not know.

Please answer the following questions regarding the Forman homicide, even if you were not assigned to this case.

Case description:

The body of Ruby Forman, an employee at the Andy Warhol Museum, was found at her home by her daughter. There was no sign of forced entry.

* 25. Ruby Forman

- ☐ lived within 2 blocks of her job at the Warhol Museum.
- ☐ died at her place of work.
- ☐ was killed by blunt force trauma to the head.
- ☐ was killed between the hours of noon and 2 PM.
- ☐ I do not know.

* 26. Beulah Tidwell was a co-worker of Ruby's who

- ☐ usually carpooled to work with Ruby.
- ☐ was thought to have motive to murder Ruby.
- ☐ found Ruby's body.
- ☐ thought that Ruby was too trusting of strangers.
- ☐ I do not know.

Please answer the following questions regarding the Kastanek homicide, even if you were not assigned to this case.

Case description:

James Kastanek's body was found by his coworker in the early morning. Robbery appeared to be the apparent motive.

* 27. James Kastanek

- ☐ had recently gotten married to Brittany Gambola.
- ☐ had tried to get Steve Gramming fired.
- ☐ was found dead on the sidewalk outside of his place of work.
- ☐ was hoping for a promotion at his job at MacDonalds.
- ☐ I do not know.

* 28. The cause of James Kastanek's death was determined to be:

- ☐ poison
- ☐ multiple gunshot wounds
- ☐ strangling
- ☐ blunt force trauma to the head
- ☐ I do not know.

* 29. Did you review the Raffield case?

- ☐ Yes
- ☐ No

Please answer the following questions regarding the Raffield homicide.

Case description:

Darleen Raffield, a venture capitalist, was found dead in her car in an alley behind a luxury apartment building.

* 30. On the day of his wife's murder, Ron Raffield claimed that

- ☐ he'd run into an old acquaintance on the bus.
- ☐ he'd been out of town on a business trip.
- ☐ he'd been tied up in a meeting all afternoon.
- ☐ he'd tried to call Darlene, but she never answered her cell.
- ☐ I do not know.

* 31. The Coroner's Report said Darleen Raffield

- ☐ was shot with a high powered rifle at the stadium.
- ☐ was pregnant with Wade McMonagle's child.
- ☐ suffered from hereditary heart disease.
- ☐ was poisoned by something in her drink.
- ☐ I do not know.

* 32. Arturo Talamo was

- ☐ having an affair with Darleen Raffield.
- ☐ investing in the NASDAQ.
- ☐ Wade McMonagle's boss.
- ☐ investing in Darleen Raffield's new company.
- ☐ I do not know.

* 33. Wayne Millican

- ☐ was a business partner of Arturo Talamo.
- ☐ was a waiter at PNC Park.
- ☐ was a maintenance man at Allegheny General Hospital.
- ☐ was not involved in this case.
- ☐ I do not know.

* 34. Darryl VanGundy

- ☐ thought Ron Raffield hired someone to kill his wife.
- ☐ was having an affair with Darleen Raffield.
- ☐ was on a business trip with Ron Raffield.
- ☐ drove the limo in which Darleen Raffield was killed.
- ☐ I do not know.

* 35. Garry Raisch

- ☐ was paid by Arturo Talamo to eavesdrop on Wade McMonagle.
- ☐ was an ex-boyfriend of Darleen Raffield from high school.
- ☐ saw Wade McMonagle give Darleen Raffield a drink.
- ☐ owned a restaurant on the North Side near the baseball stadium.
- ☐ I do not know.

* 36. Johnathan Selway

- ☐ was Darleen Raffield's good friend from college.
- ☐ ordered a home delivery dinner to be delivered at his apartment.
- ☐ was hired by Wade McMonagle to poison Darleen Raffield.
- ☐ was going to end his relationship with Darleen Raffield.
- ☐ I do not know.

* 37. Kelvin Coale was

- ☐ the driver for Wheel Deliver, the multi-restaurant delivery service.
- ☐ person that found Darleen Raffield dead in the back alley.
- ☐ the secret funder of Wade McMonagle.
- ☐ having an affair with Darleen Raffield.
- ☐ I do not know.

* 38. Ronald Raffield was

- ☐ seeking a divorce from Darleen Raffield.
- ☐ being blackmailed by Arturo Talamo.
- ☐ returning from a business trip when his wife was killed.
- ☐ hired to assassinate Wade McMonagle.
- ☐ I do not know.

* 39. Wade McMonagle

- ☐ barely escaped a murder attempt by Arturo Talamo.
- ☐ was having an affair with Darleen Raffield.
- ☐ put poison in Darleen Raffield's drink.
- ☐ was being blackmailed by Darleen Raffield.
- ☐ I do not know.

* 40. Did you think that there was a connection between any of your 3 cold cases?

- ☐ Yes
- ☐ No

* 41. Did you think that there was any connection between your cold cases and the cases from cold cases from your partner?

- ☐ Yes
- ☐ No

* 42. Did you think that there was any connection between the Raffield case and any of the cold cases?

- ☐ Yes
- ☐ No

* 43. Did you suspect that someone from the Raffield case might be the serial killer?

- ☐ Yes
- ☐ No

* 44. Did you suspect that a bus rider might be the serial killer?

- ☐ Yes
- ☐ No

* 45. Did you suspect that an Allegheny General Hospital employee might be the serial killer?

- ☐ Yes
- ☐ No

We are interested in how you worked with your partner. Please answer the following questions regarding your experience.

* 46. How well did you know your partner before the task already?

- ☐ Not at all
☐ Acquaintances
☐ Friends
☐ Close Friends
☐ Married/Domestic Partners/Dating

* 47. Think about your work with your partner today, as you answer these questions.

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
My partner and I did our jobs without getting in each other's way.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner kept me up to date about what they were doing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to discuss the cases with my partner.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner finished their tasks in time for me to do mine.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner and I often disagreed about who should be doing what task.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner and I wasted a lot of time.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner and I agreed about how to solve the case.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner treated me fairly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner had a definite sense of direction and purpose.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If given the choice, I would work with this partner again.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please answer the following questions about yourself.

48. Are you male or female?

- ☐ Male ☐ Female

49. Occupation

What is your occupation?

50. What is the highest education level you completed (or your current level if you are still in school)?

- ☐ high school
- ☐ undergraduate degree (or equivalent)
- ☐ graduate degree (or equivalent)

51. What is your Major of study?

52. How old are you?

- ☐ under 25 years old
- ☐ 25-29
- ☐ 30-34
- ☐ 35-39
- ☐ over 39 years old

53. What is your ethnicity? (check all that apply)

- ☐ African American
- ☐ American Indian or Alaskan Native
- ☐ Asian or Pacific Islander
- ☐ Hispanic or Latino
- ☐ White

Other (please specify)

54. Country of birth

In what country were you born?

55. Native Language

What is your native language?

56. Years in the United States

How many years have you been living in the United States?

57. How often do you watch TV programs like CSI, Law and Order, or Court TV?

- ☐ Never
- ☐ Once show a month or less
- ☐ One show a week or less
- ☐ 2-3 shows a week
- ☐ 4-5 shows a week
- ☐ 6+ shows a week

58. How often do you read mysteries or courtroom drama books?

- ☐ Never
- ☐ One book a year
- ☐ 2-3 books a year
- ☐ One book a month
- ☐ One book every 2 weeks
- ☐ One book a week

Thank you for completing this survey and for your participation in our experiment. Please let your experimenter know that you are finished with the experiment.

APPENDIX A.2

* 1. Subject Number

This survey will ask you some questions about your experience of participating in the Detective Experiment. It should take approximately 10 minutes to complete.

The first part of the survey contains a few brief word problems. Please record your answer to the following 3 questions in the text boxes provided.

* 2. A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?

* 3. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

* 4. In a lake there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take the patch to cover half of the lake?

The second part of this survey is to ask you about how much work you thought this experiment required.

* 5. Please select a number from 1 (LOW) to 7 (HIGH) to answer the following questions.

	(Low) 1	2	3	4	5	6	(High) 7
MENTAL DEMAND: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
TEMPORAL DEMAND: How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
EFFORT: How hard did you have to work (mentally and physically) to accomplish your level of performance?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERFORMANCE: How successful do you think you were in accomplishing the goals of the task set by the experimenter?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
FRUSTRATION: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Now we will ask you some questions about the documents in your cases. We are interested in what kinds of information stood out.

Please answer the following questions regarding the Moeller homicide, even if you were not assigned to this case.

Case description:

Madeline Moeller, an employee at Allegheny General Hospital, was found dead by her husband on the front porch of their residence.

* 6. A neighbor reported seeing the following on the day that Madeline was murdered:

- ☐ a man exiting the bus carrying a toolbox.
- ☐ a man hanging around outside Madeline's house before she got home.
- ☐ Madeline arguing loudly with her husband as they left the house for work.
- ☐ Madeline catching a ride to work with a female who stopped at her house.
- ☐ I do not know.

* 7. A list of bus riders in the Moeller case

- ☐ did not provide any useful information to solve the serial killer case.
- ☐ included the name of Wayne Millican, who was unemployed at the time.
- ☐ indicated that a business partner of Moeller's husband had been on the bus that day.
- ☐ indicated that a known sex offender, James Stuart Parsley, had been on the bus that day.
- ☐ I do not know.

Please answer the following questions regarding the Fisk homicide, even if you were not assigned to this case.

Case description:

Vicki Fisk, a student at ITT Tech, was found dead near her home by her roommate.

* 8. Vicki Fisk

- ☐ worked as a bartender in Shadyside.
- ☐ had just come to Pittsburgh from Chicago and was unemployed.
- ☐ was murdered between 10 and 10:30 PM.
- ☐ had called her boyfriend on her cell phone from the bus shortly before she was murdered.
- ☐ I do not know.

* 9. Julia Tutt

- ☐ lived with Vicki Fisk near a bus stop.
- ☐ lived with her parents but was hoping to get her own apartment soon.
- ☐ didn't like living in Pittsburgh.
- ☐ complained to friends that she didn't feel safe in Highland Park.
- ☐ I do not know.

Please answer the following questions regarding the Stokes homicide, even if you were not assigned to this case.

Case description:

William T.K. Stokes was shot twice while standing on a sidewalk. Two neighbors witnessed the crime.

* 10. T.K. Stokes

- ☐ had served time for burglary and had recently gotten out of prison.
- ☐ was about to start attending college before he was murdered.
- ☐ had been in a gang but had been trying to get out of it.
- ☐ was known to volunteer at the Boys and Girls Club.
- ☐ I do not know.

* 11. The name of T.K. Stokes's girlfriend was

- ☐ Brittany West
- ☐ Tanya Harmon
- ☐ Janice Truett
- ☐ Sherese McDonald
- ☐ I do not know.

Please answer the following questions regarding the Waring homicide, even if you were not assigned to this case.

Case description:

Myra Waring, a Mercy Hospital employee, was found by her sister with multiple wounds to the head by a blunt instrument.

* 12. Myra Waring

- ☐ was killed by her ex-husband in a fit of rage.
- ☐ was killed around 4 PM, shortly after she got off her shift at Mercy Hospital.
- ☐ was killed on the way back from a jog in her neighborhood one morning.
- ☐ was missing for several days before her body was found.
- ☐ I do not know.

* 13. Gregory Waring's alibi for his ex-wife's murder was:

- ☐ he was at work.
- ☐ he was out of town.
- ☐ he was at a dentist appointment.
- ☐ he didn't have an alibi.
- ☐ I do not know.

Please answer the following questions regarding the Forman homicide, even if you were not assigned to this case.

Case description:

The body of Ruby Forman, an employee at the Andy Warhol Museum, was found at her home by her daughter. There was no sign of forced entry.

* 14. Ruby Forman

- ☐ lived within 2 blocks of her job at the Warhol Museum.
- ☐ died at her place of work.
- ☐ was killed by blunt force trauma to the head.
- ☐ was killed between the hours of noon and 2 PM.
- ☐ I do not know.

* 15. Beulah Tidwell was a co-worker of Ruby's who

- ☐ usually carpooled to work with Ruby.
- ☐ was thought to have motive to murder Ruby.
- ☐ found Ruby's body.
- ☐ thought that Ruby was too trusting of strangers.
- ☐ I do not know.

Please answer the following questions regarding the Kastanek homicide, even if you were not assigned to this case.

Case description:

James Kastanek's body was found by his coworker in the early morning. Robbery appeared to be the apparent motive.

* 16. James Kastanek

- ☐ had recently gotten married to Brittany Gambola.
- ☐ had tried to get Steve Gramming fired.
- ☐ was found dead on the sidewalk outside of his place of work.
- ☐ was hoping for a promotion at his job at MacDonalds.
- ☐ I do not know.

* 17. The cause of James Kastanek's death was determined to be:

- ☐ poison
- ☐ multiple gunshot wounds
- ☐ strangling
- ☐ blunt force trauma to the head
- ☐ I do not know.

* 18. Did you review the Raffield case?

- ☐ Yes
- ☐ No

Please answer the following questions regarding the Raffield homicide.

Case description:

Darleen Raffield, a venture capitalist, was found dead in her car in an alley behind a luxury apartment building.

* 19. On the day of his wife's murder, Ron Raffield claimed that

- ☐ he'd run into an old acquaintance on the bus.
- ☐ he'd been out of town on a business trip.
- ☐ he'd been tied up in a meeting all afternoon.
- ☐ he'd tried to call Darlene, but she never answered her cell.
- ☐ I do not know.

* 20. The Coroner's Report said Darleen Raffield

- ☐ was shot with a high powered rifle at the stadium.
- ☐ was pregnant with Wade McMonagle's child.
- ☐ suffered from hereditary heart disease.
- ☐ was poisoned by something in her drink.
- ☐ I do not know.

* 21. Arturo Talamo was

- ☐ having an affair with Darleen Raffield.
- ☐ investing in the NASDAQ.
- ☐ Wade McMonagle's boss.
- ☐ investing in Darleen Raffield's new company.
- ☐ I do not know.

* 22. Wayne Millican

- ☐ was a business partner of Arturo Talamo.
- ☐ was a waiter at PNC Park.
- ☐ was a maintenance man at Allegheny General Hospital.
- ☐ was not involved in this case.
- ☐ I do not know.

* 23. Darryl VanGundy

- ☐ thought Ron Raffield hired someone to kill his wife.
- ☐ was having an affair with Darleen Raffield.
- ☐ was on a business trip with Ron Raffield.
- ☐ drove the limo in which Darleen Raffield was killed.
- ☐ I do not know.

* 24. Garry Raisch

- ☐ was paid by Arturo Talamo to eavesdrop on Wade McMonagle.
- ☐ was an ex-boyfriend of Darleen Raffield from high school.
- ☐ saw Wade McMonagle give Darleen Raffield a drink.
- ☐ owned a restaurant on the North Side near the baseball stadium.
- ☐ I do not know.

* 25. Johnathan Selway

- ☐ was Darleen Raffield's good friend from college.
- ☐ ordered a home delivery dinner to be delivered at his apartment.
- ☐ was hired by Wade McMonagle to poison Darleen Raffield.
- ☐ was going to end his relationship with Darleen Raffield.
- ☐ I do not know.

* 26. Kelvin Coale was

- ☐ the driver for Wheel Deliver, the multi-restaurant delivery service.
- ☐ person that found Darleen Raffield dead in the back alley.
- ☐ the secret funder of Wade McMonagle.
- ☐ having an affair with Darleen Raffield.
- ☐ I do not know.

* 27. Ronald Raffield was

- ☐ seeking a divorce from Darleen Raffield.
- ☐ being blackmailed by Arturo Talamo.
- ☐ returning from a business trip when his wife was killed.
- ☐ hired to assassinate Wade McMonagle.
- ☐ I do not know.

* 28. Wade McMonagle

- ☐ barely escaped a murder attempt by Arturo Talamo.
- ☐ was having an affair with Darleen Raffield.
- ☐ put poison in Darleen Raffield's drink.
- ☐ was being blackmailed by Darleen Raffield.
- ☐ I do not know.

* 29. Did you think that there was a connection between your 4 cold cases?

- ☐ Yes
- ☐ No

* 30. Did you think that there was any connection between your cold cases and the cases from cold cases from your partner?

- ☐ Yes
- ☐ No

* 31. Did you think that there was any connection between the Raffield case and any of the cold cases?

- ☐ Yes
- ☐ No

* 32. Did you suspect that someone from the Raffield case might be the serial killer?

- ☐ Yes
- ☐ No

* 33. Did you suspect that a bus rider might be the serial killer?

- ☐ Yes
- ☐ No

* 34. Did you suspect that an Allegheny General Hospital employee might be the serial killer?

☐ Yes

☐ No

We are interested in how you worked with your partner. Please answer the following questions regarding your experience.

* 35. How well did you know your partner before the task already?

☐ Not at all

☐ Acquaintances

☐ Friends

☐ Close Friends

☐ Married/Domestic Partners/Dating

* 36. Think about your work with your partner today, as you answer these questions.

	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
My partner and I did our jobs without getting in each other's way.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner kept me up to date about what they were doing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to discuss the cases with my partner.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner finished their tasks in time for me to do mine.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner and I often disagreed about who should be doing what task.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner and I wasted a lot of time.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner and I agreed about how to solve the case.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner treated me fairly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My partner had a definite sense of direction and purpose.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If given the choice, I would work with this partner again.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* 37. If your partner shared a clue with you on the phone, please rate how useful it was?

- ☐ Not very useful
- ☐ Not useful
- ☐ was ok
- ☐ Was useful
- ☐ Was very useful
- ☐ Partner did not share any clue

* 38. If your partner shared a clue with you on the phone, please rate the timing?

- ☐ very bad
- ☐ bad
- ☐ was ok
- ☐ good
- ☐ very good
- ☐ partner did not share any clue

Please answer the following questions about yourself.

* 39. Are you male or female?

- ☐ Male ☐ Female
- ☐ Other (please specify)

* 40. Occupation

What is your occupation?

* 41. What is the highest education level you completed (or your current level if you are still in school)?

- ☐ high school
- ☐ undergraduate degree (or equivalent)
- ☐ graduate degree (or equivalent)

* 42. What is your Major of study?

* 43. How old are you?

- ☐ under 25 years old
- ☐ 25-29
- ☐ 30-34
- ☐ 35-39
- ☐ over 39 years old

* 44. What is your ethnicity? (check all that apply)

- ☐ African American
- ☐ American Indian or Alaskan Native
- ☐ Asian or Pacific Islander
- ☐ Hispanic or Latino
- ☐ White

Other (please specify)

* 45. Country of birth

In what country were you born?

* 46. Native Language

What is your native language?

* 47. Years in the United States

How many years have you been living in the United States?

* 48. How often do you watch TV programs like CSI, Law and Order, or Court TV?

- ☐ Never
- ☐ Once show a month or less
- ☐ One show a week or less
- ☐ 2-3 shows a week
- ☐ 4-5 shows a week
- ☐ 6+ shows a week

* 49. How often do you read mysteries or courtroom drama books?

- ☐ Never
- ☐ One book a year
- ☐ 2-3 books a year
- ☐ One book a month
- ☐ One book every 2 weeks
- ☐ One book a week

* 50. I see myself as someone who

	Disagree				Agree
is reserved	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
is generally trusting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
tends to be lazy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
is relaxed, handles stress well	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
has few artistic interests	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
is outgoing, sociable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
tends to find faults with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
does a thorough job	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
gets nervous easily	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
has an active imagination	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Thank you for completing this survey and for your participation in our experiment. Please let your experimenter know that you are finished with the experiment.